DEC 5 1944
GENERAL LIBRARY
UNIV. OF MICH.

OCTOBER, 1944

Journal of Educational Research

Registered June 4, 1929

CONTENTS

		PAGE
Difficulty Ratings of Addition Facts as Related to Learning Method	*Esther J. Swenson*	81
The Relationship of Reading Skills to the Ability to Solve Arithmetic Problems	*John P. Treacy*	86
The Effect of Instruction in Mathematical Vocabulary Upon Problem Solving in Arithmetic	*Harry C. Johnson*	97
Factors Associated with Successful Achievement in Problem Solving in Sixth Grade Arithmetic	*Carl W. Hansen*	111
Basic Considerations for Profitable Research in Arithmetic	*Guy M. Wilson*	119
The Mathematical Vocabulary Used in Secondary School Textbooks of Science	*Francis D. Curtis*	124
Factors Which Affect the Salary of High School Teachers of Commercial Subjects in Missouri	*Harry Pelle Hartkemeier*	132
Some Major Problems in Educational Measurement	*Benjamin S. Bloom*	139
Directional Tests for Educational Guidance	*Forrest H. Kirkpatrick*	143
Editorials:		
The Synthesis of Scientific Studies	*Leo J. Brueckner*	146
The Increase in Precision in Educational and Psychological Experimentation Through Statistical Controls	*Palmer O. Johnson*	149
Research Abstracts and Bibliographies	*A. S. Barr*	152
Research News and Communications	*Carter V. Good*	157

THE JOURNAL OF EDUCATIONAL RESEARCH is indexed in the Education Index

Edited and published by A. S. Barr, Professor of Education, University of Wisconsin, Madison, Wisconsin

Entered as second-class matter at the post office of Madison, Wisconsin under the act of March 3, 1879

JOURNAL OF EDUCATIONAL RESEARCH

EDITORIAL BOARD

A. S. BARR, *Chairman, Professor of Education, Madison 6, Wisconsin*

E. J. ASHBAUGH, *Dean College of Education, Miami University, Oxford, Ohio.* Educational Administration, and Teacher Training.

LEO J. BRUECKNER, *Professor of Education, University of Minnesota, Minneapolis, Minnesota.* Elementary Education.

HARL R. DOUGLASS, *Director of the School of Education, University of Colorado, Boulder, Colorado.* Secondary Education.

CARTER V. GOOD, *Professor of Education, University of Cincinnati, Cincinnati, Ohio.* Teaching, Learning, and Supervision.

A. K. LOOMIS, *Director, School of Education, University of Denver, Denver, Colorado.* Curriculum Construction.

DOUGLAS E. SCATES, *Associate Professor of Education, Duke University, Durham, N. C.* Field Research.

PERCIVAL M. SYMONDS, *Professor of Education, Teachers College, Columbia University, New York City.* Mental Hygiene and Pupil Adjustment.

CLIFFORD WOODY, *Director, Bureau of Educational Reference and Research and Professor of Education, University of Michigan, Ann Arbor, Michigan.* Measurements, Statistics, and Research Techniques.

To save time, manuscripts may be sent directly to members of the editorial board

ASSOCIATE EDITORS

FRED C. AYER, *Professor of Educational Administration, University of Texas, Austin, Texas.*

HARRY J. BAKER, *Director, Psychological Clinic, Detroit Public Schools, Detroit 1, Michigan.*

EMMETT A. BETTS, *Research Professor and Director of the Reading Clinic, Pennsylvania State College, State College, Pennsylvania.*

PHILIP A. BOYER, *Director of Educational Research School District of Philadelphia, Philadelphia, Pennsylvania.*

FOWLER D. BROOKS, *Head of Department of Psychology, DePauw University, Greencastle, Indiana.*

WILLIAM A. BROWNELL, *Professor of Educational Psychology, Duke University, Durham, North Carolina.*

B. R. BUCKINGHAM, *Directing Editor of Elementary School Textbooks, Ginn and Company, Boston, Massachusetts.*

W. W. CHARTERS, *Director of Research, Stephens College, Columbia, Missouri.*

VIRGIL E. DICKSON, *Superintendent of Schools, Berkeley Public Schools, Berkeley, California.*

EDGAR A. DOLL, *Director, Department of Research, Training School, Vineland, New Jersey.*

WILLIAM S. GRAY, *Executive Secretary of the Committee on the Preparation of Teachers, University of Chicago, Chicago, Illinois.*

CHARLES E. GREENE, *Superintendent of Schools, Denver Public Schools, Denver, Colorado.*

V. A. C. HENMON, *Professor of Psychology, University of Wisconsin, Madison, Wisconsin.*

KARL J. HOLZINGER, *Professor of Education, University of Chicago, Chicago, Illinois.*

ERNEST HORN, *Director, University Elementary School and Professor of Education, State University of Iowa, Iowa City, Iowa.*

EARL HUDELSON, *Dean, College of Education and Professor of Education, West Virginia University, Morgantown, West Virginia.*

FREDERICK J. KELLY, *Chief, Division of Higher Education, Office of Education, Federal Security Agency, Washington, D. C.*

CHARLES W. KNUDSEN, *Professor of Education, George Peabody College for Teachers, Nashville, Tennessee.*

J. MURRAY LEE, *Dean, School of Education, State College of Washington, Pullman, Washington.*

D. WELTY LEFEVER, *Professor of Education, University of Southern California, Los Angeles, California.*

ELISE H. MARTENS, *Senior Specialist in the Education of Exceptional Children, Office of Education, Federal Security Agency, Washington, D. C.*

WILLIAM A. McCALL, *Professor of Education, Teachers College, Columbia University, New York City.*

W. J. OSBURN, *Professor of Education, University of Washington, Seattle, Washington.*

JESSE B. SEARS, *Professor of Education, Stanford University, California.*

MANDEL SHERMAN, *Director of Orthogenic School, Professor of Educational Psychology, University of Chicago, Chicago, Illinois.*

FRANCIS T. SPAULDING, *Professor of Education, and Dean of Graduate School of Education, Harvard University, Cambridge, Massachusetts.*

JOHN L. STENQUIST, *Director, Bureau of Educational Research, Department of Education, Baltimore, Maryland; and Lecturer, University of Maryland.*

W. W. THEISEN, *Assistant Superintendent of Schools, Milwaukee, Wisconsin.*

M. R. TRABUE, *Dean, School of Education, Pennsylvania State College, State College, Pennsylvania.*

CARLETON W. WASHBURNE, *Superintendent of Schools, Winnetka Public Schools, Winnetka, Illinois.*

ELIZABETH L. WOODS, *Supervisor, Educational Research and Guidance, Los Angeles City Schools, Los Angeles, California.*

J. WAYNE WRIGHTSTONE, *Assistant Director, Bureau of Reference, Research, and Statistics, Board of Education of the City of New York, Brooklyn, N. Y.*

G. A. YOAKAM, *Professor of Education and Director, Courses in Elementary Education, University of Pittsburgh, Pittsburgh, Pennsylvania.*

The JOURNAL OF EDUCATIONAL RESEARCH is edited and published at Madison, Wisconsin, monthly except June, July, and August by A. S. Barr. The subscription price is $3.50; $4.00 in Canada and in other foreign countries. Single copies 50¢.

Address all communications to A. S. Barr, Department of Education, University of Wisconsin, Madison 6, Wisconsin.

Democrat Printing Company, Madison, Wisconsin

Copyright 1944 by A. S. Barr, Madison 6, Wisconsin

JOURNAL *of* EDUCATIONAL RESEARCH

Volume XXXVIII OCTOBER, 1944 Number 2

DIFFICULTY RATINGS OF ADDITION FACTS AS RELATED TO LEARNING METHOD

ESTHER J. SWENSON
Ball State Teachers College

Editor's Note: The comparative difficulty of learning seems to be at least, in part, a function of teaching method. The author presents data that support this thought in the learning and teaching of arithmetic.

TOO often in the field of education have the results of careful scientific research been ignored insofar as their application in practical teaching and learning situations is concerned. A notable exception is afforded by the use which has been made of research studies on the difficulty of learning so-called arithmetic "facts". Writers of textbooks and courses of study as well as classroom teachers have used the results of such research to determine the order in which the said "facts" should be presented, the assumption being that the optimum learning order should correspond to the "difficulty of learning" order determined by the research studies. Many a sales argument has been based upon this assumption.

Of course, none of these difficulty studies has ever agreed too closely with any other. Such lack of agreement indicates a certain degree of unreliability in the ratings derived from either or both of the studies being compared. Common practice has been either (1) to use a composite rating based on two or more studies or (2) to use the ratings derived from that study which appealed to the user as being more dependable.

Since most of the research on relative difficulty of arithmetic facts has been based on repetitive methods of learning, the question arises as to whether or not such results are valid when applied in learning situations not so definitely drill-centered. For example, can the order of difficulty of the addition facts as derived from the use of drill-type learning procedures be depended upon as being the correct order of difficulty for learning the

same facts by some other learning method? Or is the order of difficulty a function of learning method? The study here reported was directed toward securing a reply to this question.

The study in question was a subsidiary investigation in connection with an extensive study[1] of forgetting and transfer of training, using the 100 addition facts as the learning material for 332 second-grade children. Individual subjects participating in the experiment varied in C. A. from 70 to 130 months, in M. A. from 71 to 114 months, and in I. Q. from 74 to 146. By a stratified randomization procedure, each second-grade class was assigned to one of three different instructional methods (designated as D, X, and G).

The near-equivalence of the three large method groups is indicated by the mean M. A.'s (D, 94.29; X, 94.22; and G, 94.44) and the mean numbers of addition facts known preceding instruction (D, 24.52; X, 24.26; and G, 24.00). By analysis of variance it was shown that the D, X, and G groups were not significantly different in either M. A. or original knowledge of the addition facts. Instruction on the addition facts continued for a period of 15 weeks, with 25 minutes of instruction each school day in all groups.

Since the purpose of this article is not to compare the relative merits of the different instructional procedures but only to point out the relationship between method of learning and difficulty ratings of facts, the three learning methods will not be described in detail. The D method was based, presumably, upon drill learning theory, with stress upon varied repetition of addition facts in miscellaneous order. The facts were presented to this group in order of difficulty as found by combining the results of three "difficulty studies".[2] The G method was based upon the generalization or meaning theory of learning arithmetic, with stress upon discovery of relationships among facts and organization of those facts around number relationships. Order of presentation was determined by a set of number generalizations

[1] A complete report of the study is to be found in the writer's doctoral dissertation. Swenson, Esther J. *Generalization and Organization as Factors in Transfer and Retroactive Inhibition.* University of Minnesota. Doctor's Thesis, 1941.

A summary of the study has also been published under the same title in the *Proceedings of the Indiana Academy of Science,* pp. 248–255, 1942.

[2] Knight, F. B., and Behrens, M. S., *The Learning of the One Hundred Addition Combinations and the One Hundred Subtraction Combinations.* New York, 1928.

Thiele, C. L. The contribution of generalization to the learning of the addition facts. *Teachers College Contributions to Education.* No. 763, 1938.

Wheeler, Lester R., A comparative study of the difficulty of the 100 addition combinations. *Journal of Genetic Psychology,* 54:295–312, 1939.

around which facts were grouped (e.g., "adding 1", "adding 0", "adding near-doubles", etc.). The X method was an intermediate method designed to correspond as nearly as possible to "common practice" in the teaching of arithmetic. There was much emphasis upon drill, but it was accompanied by concrete presentation of each addition fact and the facts were presented in an organized pattern, that of "size-of-sum".

Timed tests were given at the beginning and end of the experiment. Combinations were presented by a flash-card technique at the rate of one every 4 seconds. A brief pause was allowed after every tenth fact. Answers only were written by the children on answer sheets. The numbers of correct, omitted, and wrong responses to each test item were tabulated. The tabulated data were handled separately for each method, and percentages of difficulty (number right ÷ number responding) computed for each fact.

Such percentage ratings were made for each fact within each method group. The 100 facts were assigned ranks from 1 (easiest) to 100 (hardest). For instance, the difficulty ranks of 1 + 1 on the initial test were: for the X group, 2; for the D group, 1; and for the G group, 1. As another example, 6 + 7 had these ranks on the initial test: X, 89; D, 87; and G, 88. Coefficients of correlation were computed to indicate the relationship between difficulty ranks within the different method groups—that is, between X and D ratings, between D and G ratings, and between X and G ratings. The formula used for these correlations was:

$$\frac{\text{Sum XY} - \dfrac{(\text{Sum X})\ (\text{Sum Y})}{\text{N}}}{\sqrt{\left(\text{Sum X}^2 - \dfrac{(\text{Sum X})^2}{\text{N}}\right)\left(\text{Sum Y}^2 - \dfrac{(\text{Sum Y})^2}{\text{N}}\right)}}$$

Table 1 presents the correlation coefficients for both the initial and final tests. The *r* for the relation between X and D ratings on the first test was .969; for the relation between G and X, .971; and for the relation between G and D, .967. These highly significant *r's* would seem to indicate very high agreement in difficulty ratings from one group to another at the beginning of the experiment, before the children had had any definite instruction on the addition facts.

The question to be answered was whether the different instructional procedures would result in appreciable differences in difficulty ratings of the addition facts. To answer that question, *r's* were computed for ranks on the

final test in the same manner as for the initial test. The r for X and D was .910, for G and X, .88; and for G and D, .934 (see Table 1). These r's, like those on the first test, are highly significant and indicate a close relationship between difficulty ratings at the end of the study in spite of the use of different methods of instruction in the different groups.

TABLE I

SIGNIFICANCE OF DIFFERENCE IN CORRELATIONS OF DIFFICULTY RATINGS BEFORE AND AFTER INSTRUCTION BY DIFFERENT METHODS

Ratings Compared	r	z_r	Difference	N	σ_z	$\sigma_{diff.}$	Diff. / σ diff.	P
X_I D_I	.969	2.0756	0.5481	100	.1015	.1435	3.82	.0001
X_F D_F	.910	1.5275		100	.1015			
G_I X_I	.971	2.1095	0.7337	100	.1015	.1435	5.11	beyond .0001
G_F X_F	.880	1.3758		100	.1015			
G_I D_I	.967	2.0439	0.3551	100	.1015	.1435	2.47	.0135
G_F D_F	.934	1.6888		100	.1015			

I = Initial test ratings. F = Final test ratings.

However, the fact that there remains a highly significant relationship between the difficulty ratings in different method groups at the end of the study does not answer the question as to whether or not a significant difference in relationships occurred during the learning period. To find this out, the significance of the difference between the r determined at the beginning of the study and the corresponding r determined at the end of the study must be considered. The significance of the differences between the initial test r's and the final test r's was determined by the use of the following formulas.[3]

$$t = \frac{Z_{r_1} - Z_{r_2}}{S.E._{diff.}} \qquad S.E._{diff.} \sqrt{S.E._{z_1}^2 + S.E._{z_2}^2}$$

$$S.E._z = \frac{1}{\sqrt{N-3}} \qquad z_r = \frac{1}{2}\left[\log_e (1+r) - \log_e (1-r)\right]$$

Table 1 makes it quite apparent that the different methods of instruction (including their different orders of presentation of number facts) were

[3] The values of z_r were read from Treloar's tables of z_r in Treloar. *Elements of Statistical Reasoning*. New York: John Wiley and Sons. 1939, p. 244.

accompanied by a significant change in correlation coefficients. For instance, the difference between X and D *r* on the first test and the X and D *r* on the final test was highly significant. The t-value of 3.82 reaches the .0001 level of probability. The difference between initial and final *r's* for the relationship between G and X ratings was even beyond the .0001 level of probability ($t = 5.11$). The G and D difficulty correlations for initial and final tests were also significantly different ($t = 2.47$) at the .0135 level of probability.

In every case the significant difference between *r's* was caused by a *reduction* in the size of the correlation coefficient during instruction. It seems that the different teaching methods did result in enough changes in relative difficulty of addition facts from one method to another to cause significant decreases in the inter-method difficulty correlations.

An inspection of the data from this experiment leads one to the conclusion that instruction in the 100 addition facts by methods varying chiefly in their degree of emphasis upon organization and generalization seemed to result in significantly lower correlations of inter-method difficulty ratings for the addition facts. A logical implication is that research workers in the field of arithmetic should plan their research and interpret their results in terms of the type of learning situation in which children are doing their learning.

Research which aims toward establishing the difficulty of arithmetic skills and processes should probably do so in terms of a clearly defined teaching and learning method. It may well be that many of the laboratory difficulty studies are not so widely applicable under varied classroom situations as has been supposed. If such studies were repeated under other conditions than those under which they were originally done, and results compared, quite different results might emerge. The implication of the present study that comparative difficulty of addition facts may be to some extent a function of learning method is worthy of further investigation.

THE RELATIONSHIP OF READING SKILLS TO THE ABILITY TO SOLVE ARITHMETIC PROBLEMS

JOHN P. TREACY
Graduate School, Marquette University

Editor's Note: Reading has been shown to be a function of achievement in the content subjects. The author explores this relationship in the teaching of arithmetic.

THE PROBLEM

THE purpose of this study was to determine the relationship of certain reading skills to ability to solve verbal problems in arithmetic. General questions which grew out of this problem are:

1. Is general reading level significantly related to ability to solve problems in arithmetic?

2. Are certain specific reading skills significantly related to ability to solve problems in arithmetic?

DEFINITION OF TERMS

The interpretation of "reading skills" is limited in this study to the abilities measured by 15 specific tests. "Problems" is used here to denote verbal arithmetic problems which involve the use of computations. "Problem solving ability" refers to the traits measured by two tests of arithmetic reasoning. "Relationship" is interpreted as concomitant occurrence, not as causal relationship.

THE PLACE OF PROBLEM SOLVING IN ARITHMETIC

According to an older view, the ultimate objective of arithmetic instruction was the solution of practical problems of life.[1] Recently broader and deeper values are claimed for arithmetic in general, and for problem solving in particular. Arithmetic instruction has now come to be thought of in terms of its computational, informational, sociological, and psychological functions.[2] Because of this shift in objectives, arithmetic is considered a social as well as a practical subject.[3] Ability to solve practical problems is still among arithmetic objectives, but it is only one objective among many.

PROBLEM SOLVING IN ARITHMETIC AND IN LIFE

Dewey identifies a problem with any situation involving perplexity, confusion or doubt; he thinks of a problem as containing elements which

are unexpected, queer, strange, funny, or disconcerting. To Dewey a problem is the starting point of reflective thinking.[4]

Problems in arithmetic textbooks and in tests sometimes are not really problems at all in this sense, but are merely practice exercises for operations which are being habituated.[5] They often do not involve the evaluative and emotional elements which characterize the situations in which life problems are found.[6] The newer textbooks do tend to group problems about significant life-like situations such as installment buying, insurance, and the like.[7]

Problem solving is a typical form of thinking in life situations. The approaches to problem solving are very similar, regardless of what the area of experience happens to be. Because of this, solving problems in arithmetic can be a positive or a negative force in helping pupils in the solution of other life problems. Insofar as factors which influence the solving of verbal problems in arithmetic are known and utilized intelligently in teaching, the pupils' opportunities to sense and to practice the characteristic elements in problem solving (formulation, solution, data, approximation, function, operation, proof, and symbolism) are thereby enhanced. And, correspondingly, insofar as pupils are deprived of experience in problem solving through interference of such factors as reading weaknesses, their possibilities for mastering these elements are thereby hindered.

RESEARCH ON FACTORS RELATED TO SUCCESS IN PROBLEM SOLVING

Factors which have been studied in attempts to analyze difficulties in problem solving include: size of numbers; physical defects; terms used, especially those involving quantitative relations; sex of pupils; methods used; systematical arrangement of work; nature of supervision; general vocabulary; grasp of essential facts, i.e., tables of weights and measures; the habit of checking results; presence of non-essential elements; background of experience; reading level; individual preferences; energy, effort; command of fundamentals; general intelligence; and ability to do reflective thinking.[8]

Some of the studies on factors in problem solving were based upon subjective analyses which could contribute little of a definite nature. Many studies were based on correlations, with no attempt made to control some factors while measuring the effects of others. Most studies did not show whether or not the correlation found could have been due to chance, or whether or not apparent differences in sizes of correlations were statistically significant. In most of the studies using a correlation technique, positive

correlations were interpreted as representing a causal relationship, although there was no follow-up study to check on this assumption. A number of studies of the factors related to problem solving included reading among a number of other items. In most of these instances reading was not defined, and was not broken down into its component skills. Because of these and other limitations of existing research on factors related to problem solving, broad generalizations regarding related factors are hardly in order. The research which has been conducted may justify certain tentative conclusions:

1. General intelligence, vocabulary, use of quantitative terms, and computational ability are positively related to problem solving.

2. Systematic training in a definite procedure for attacking problems results in higher scores on problem tests. The specific procedure which is most effective is not clear, and may be conditioned by such factors as the zeal of the teacher, the age of the pupils, and by similar conditions.

3. Studies of the relationship of reading to problem solving ability are somewhat contradictory. A possible explanation of these discrepancies is that reading was not interpreted or measured in the same way in the various studies. If different reading skills are not equally important for success in problem solving, and if the reading skills measured varied among the studies, it was to be expected that there would be discrepancies in the findings regarding the relationship of reading to problem solving. The findings of the present study may help to explain certain findings which have been reported. The weight of the evidence indicates that there is a positive relationship between reading as it has been measured and problem solving ability.

PROBLEM SOLVING AND SPECIFIC READING SKILLS

The need for studies of the relationship of specific reading skills to problem solving ability was anticipated a number of times by Monroe and Englehart.[9] Illustrations of this are:

> Reading ability seems to be an important factor in the ability to respond to verbal problems, but the precise nature of its function has not been ascertained.[10]
>
> There is also need for more information about the function of reading in pupil responses to verbal problems and the relation of the form and vocabulary of problem statements to these responses.[11]

Wilson recently has expressed even more clearly the nature of the problem:

There is an accumulation of evidence which indicates that the reading of verbal problems calls for some specific reading skills as well as for an acquaintance with the vocabulary and conventions employed in problem statements.

The question of the nature of the reading instruction that should be given has received only limited attention and further research is needed before any conclusion can be stated.[12]

PROCEDURES IN PRESENT STUDY

Eighteen items of information were collected for each of 244 7B pupils in two Milwaukee junior high schools through administering the following tests:

Tests of problem solving:

Analytical Scales of Attainment: "Problems."[14]
Public School Achievement Tests: "Arithmetic Reasoning."[14]

Tests of mental ability:

Otis Quick-Scoring Mental Test, Beta, Form A[15]

Tests of reading ability:

Gates Silent Reading Tests[16]
- Type B, Prediction of Outcomes
- Type C, Understanding of Directions

Analytical Scales of Attainment[9]
- Quantitative Relationships
- Arithmetic Vocabulary

Diagnostic Examination of Silent Reading Abilities[17]
- Rate of Comprehension
- Perception of Relationships
- Vocabulary in Context
- Vocabulary, Isolated Words
- General Information
- Grasp of General Thought
- Retention of Details
- Interpretation of Content
- Integration of Dispersed Ideas
- Drawing of Inferences
- General Reading Level

These tests were administered under the supervision of the writer during the period from May 20 to June 8, 1940. A total of eight sittings, involving approximately 375 minutes, were required for this testing. The

tests were scored through the cooperation of the W.P.A. Testing Project of the Milwaukee Public Schools.

The criterion for ability in problem solving ability was the average performance on two standardized tests. Since the units of measurement on these two tests were not equivalent, the raw scores could not be averaged. In order to make it possible to combine the measures of problem solving ability, the raw scores were turned into T-scores.[18] In the resulting derived scores, the unit of measurement is one-tenth of a standard deviation. In order to avoid certain statistical complications, the raw scores of reading tests were also turned into T-Scores.

The reliability of each of these tests for the group measured was determined by selecting at random[19] 50 test papers and applying the Hoyt[20] method of estimating reliability coefficients. Detailed results are reported in Table I. The reliability coefficients found for the reading tests varied from .47 in Quantitative Relationships to .97 in Vocabulary, Isolated Words; the median reliability for the reading tests was .78. All coefficients except one were above .70, eight were above .80, and two were above .90. These coefficients may be an underestimate of the actual reliability. Other evidences of reliability and validity are reported in the respective test manuals.

In order to combine the 7B groups from the two schools, it had to be shown that in problem solving they were homogeneous both as to means and as to variability. Homogeneity of variability was determined by dividing the larger variance by the smaller variance and applying Snedecor's Table of "F"[21] values. Homogeneity of means was determined by the "t" test of significance.[22]

The 80 pupils having the highest combined T-Score on the tests of problem solving were designated as "good achievers", and the 80 pupils having the lowest combined T-Score were designated as "poor achievers". Through use of the "F" test and of the "t" test it was demonstrated: (1) that the good achievers in the two classes were homogeneous as to mean achievement and as to variability in problem solving; (2) that the poor achievers in the two groups were homogeneous as to mean achievement and as to variability in problem solving; (3) and that the good achievers as to group and the poor achievers as a group were homogeneous as to variability but heterogeneous as to mean achievement in problem solving.

TABLE I

RELIABILITY AND CERTAIN VALUES USED IN COMPUTING RELIABILITY FOR ONE TEST OF INTELLIGENCE, TWO TESTS OF PROBLEM SOLVING, AND FOURTEEN READING SKILLS

Test	N	k	T	S_a	S_1	r_{tt}
1. Otis Test (Intelligence)	80	50	2614	140674	100032	.86
2. Arithmetic Problems*	25	50	710	11621	22806	.87
3. Arithmetic Reasoning**	48	50	1409	41733	53323	.75
4. Quantitative Relationships*	40	50	925	17774	28397	.47
5. Arithmetic Vocabulary*	40	50	1138	26982	39566	.70
6. Prediction of Outcomes (Gates, B)	24	50	866	15648	37502	.86
7. Understanding Directions (Gates, C)	24	50	815	14005	31579	.78
8. Perception of Relationships***	37	50	1251	32967	47697	.85
9. Vocabulary in Context***	37	50	1258	33800	46616	.87
10. Vocabulary, Isolated Words***	37	50	1198	32967	47697	.97
11. General Information***	37	50	1136	27698	39788	.84
12. Central Thought***	20	50	580	7465	18118	.74
13. Retention of Details***	20	50	568	7210	17202	.74
14. Interpretation***	20	50	581	7375	18659	.70
15. Integration of Dispersed Ideas***	20	50	591	7757	18365	.74
16. Drawing Inferences***	20	50	592	7760	18408	.75
17. Reading Level***	100	50	2912	181420	90752	.92

* Analytical Scales of Attainment.
** Public School Achievement Tests.
*** Diagnostic Examination of Silent Reading Abilities.
The Reliability of the Rate of Comprehension Test could not be measured because this test was a part of the material returned with the summary charts to the cooperating schools.

The good and poor achievers were then compared on each of 15 reading skills by means of the "t" test of significance. This technique, however, had a serious limitation, for it left uncontrolled the effects of chronological age and of mental age. Chronological age was found to correlate — .299 with problem solving ability, and mental age correlated .758 with this ability. Both of these correlations were found to be significant at the one per cent level by use of Fisher's formula for the significance of correlations.[23] Then the Johnson–Neyman technique of statistical analysis[24] was used, with mental age and chronological age statistically controlled. For both the "t" test and the Johnson–Neyman technique the null hypothesis tested was that the difference in means in the various reading skills between good and poor achievers in problem solving is zero. The facts of observation were then allowed to refute this hypothesis.

FINDINGS

1. Application of the "t" test of significance showed good achievers in problem solving to be superior to poor achievers in problem solving at the one per cent level of significance in all of the reading skills studied and in

mental age. The "t" values for these varied from 3.19 to 12.59, with a median value of 8.48. Poor achievers were older chronologically than good achievers at the five per cent level of significance.

2. A comparison of good and poor achievers in problem solving in arithmetic on 15 reading skills by means of the Johnson-Neyman technique, with mental age and chronological age statistically controlled, resulted in the following findings:

(a) Good achievers were found to be better than poor achievers at the one per cent level of significance in Quantitative Relationships, Perception of Relationships, Vocabulary in Context, and Integration of Dispersed Ideas.

(b) Good achievers were found to be better than poor achievers at the five per cent level of significance in Arithmetic Vocabulary, Vocabulary (Isolated Words), Retention of Clearly Stated Details, Drawing of Inferences from Context, and Reading Level.

(c) No significant differences were found between good and poor achievers in the Prediction of Outcomes, Understanding of Precise Directions, Rate of Comprehension, General Information, Grasp of Central Thought, and Interpretation of Content.

(d) In each instance in which a significant difference in reading skill was found between good and poor achievers in problem solving by means of the Johnson-Neyman technique, the region of significance was restricted to only a definite part of the population. The mental ages varied from 118 months to 224 months, the entire scale used in the present analysis. The chronological age varied from 128 months to 168 months. Two of the regions of significance were in the form of a hyperbola, seven in the form of an ellipse. The detailed findings are reported in Table II.

EDUCATIONAL IMPLICATIONS OF FINDINGS

1. The results of this study indicate that in studying the causal relationships between reading and problem solving in arithmetic, and in planning procedures for helping pupils through reading instruction to improve their ability in problem solving, reading should be regarded as a composite of specific skills rather than as a generalized ability. The list of reading skills which should be considered, in addition to the ones studied here, is a problem for further research.

TABLE II

SUMMARY OF FINDINGS: JOHNSON–NEYMAN ANALYSES

	Reading Skill	Region of Significance		
		Mental Age (Months)	Chronological Age (Months)	Shape*
Significant at One Per Cent	Quantitative Relationships	134–190	137–163	E
	Perception of Relations**	118–190 (202–224)	128–161 (Above 173)	H
	Vocabulary in Context	134–186	130–156	E
	Integration of Dispersed Ideas	156–224	128–154	E
Significant at Five Per Cent Level	Arithmetic Vocabulary**	182–224 (Less than 118)	128–158 (Above 158)	H
	Vocabulary, Isolated Words	118–161	128–152	E
	Clearly Stated Details	142–190	128–143	E
	Drawing Inferences	154–176	138–148	E
	Reading Level	130–196	128–152	E
No Region of Significance	Predicting Outcomes			
	Precise Directions			
	Rate of Comprehension			
	General Information			
	Central Thought			
	Interpretation of Content			

* E—Indicates that the region of significance is an ellipse; H—Indicates that the region is a hyperbola.

** Figures in parentheses represent an imaginary region favorable to poor achievers, but not represented by the present sample.

2. Teachers of arithmetic should consider the reading skills which are significantly related to success in this subject, especially in their diagnostic and remedial work. Tests covering reading skills which are related to problem solving might well precede some exercises in problem solving. Remedial teaching should follow where weaknesses in reading skills related to success in problem solving are discovered.

3. When carrying on diagnostic and remedial teaching in reading, in an attempt to improve problem solving in arithmetic, attention should be given to the mental ages and the chronological ages for which the relationship between reading skills and problem solving has been found significant.

4. Four of the reading skills on which good and poor achievers in problem solving differed significantly were associated in one way or another with vocabulary (Quantitative Relationships; Vocabulary in Context; Vocabulary, Isolated Words; and Arithmetic Vocabulary). This fact suggests

the need for stressing the meaning of terms, general and mathematical, as an approach to improving pupils' ability in problem solving.

5. Three of the reading skills on which good and poor achievers differed significantly (Perception of Relationships; Integration of Dispersed Ideas; and Drawing Inferences from Content) require that the pupil carefully associate, immediately and remotely, the elements of the reading material which he reads. This finding may be one explanation of the fact that Rate of Comprehension was not found to be a significant factor in problem solving. It also suggests that help given to a pupil which enriches the meanings of what he reads may have a beneficial effect on his problem solving ability.

6. Good and poor achievers differed significantly on Retention of Clearly Stated Details. This suggests the possibility that pupils may be troubled in problem solving by *inability to retain* significant facts as well as by inability to understand what is read, and that instruction should be designed accordingly.

7. In the review of the literature regarding the relationship of reading and problem solving in arithmetic some discrepancies in findings were found. This situation may be explained in part by the fact that some studies were based upon measures of certain reading skills, and some studies on other reading skills. In the light of the present findings, such a discrepancy would have been expected. For example, some of the studies employed reading tests in which rate of comprehension was an element; the present study did not find that skill to be significant. Other studies used tests of general reading level; the present study found reading level to be significant only at the five per cent level. Still other studies reported used reading tests which measure reading skills of unequal significance for success in problem solving. *These facts confirm an earlier conclusion, that research on the relationships of reading and problem solving in arithmetic should be in terms of general reading ability.*

CAUTIONS REGARDING THE INTERPRETATION OF RESULTS

Certain limitations must be recognized when interpreting the results of this study: (1) "Relationship" as used here refers only to concomitant occurrence; it does not necessarily imply causal relationship. (2) "Mental Age," "problem solving ability," and "reading skills" are limited here to the traits measured by the tests used in ascertaining them. (3) In some of

the analyses the expression "significant at the five per cent level" is used; it may be questionable whether or not differences at the five per cent level should be considered as significant.

SOME SUGGESTED PROBLEMS FOR FURTHER RESEARCH

1. A study of the relationship of other reading skills to ability to solve problems in arithmetic.

2. Studies similar to the present one, but using samples from different populations.

3. Controlled studies of the influence of improvement in certain reading skills on progress in problem solving in arithmetic.

4. Studies of the inter-relationships of various skills which are measured by reading tests.

5. Studies of the reading skills needed to solve arithmetic problems which occur in the lives of adults.

6. A comparison of the reading skills needed to solve problems in the older and in the newer arithmetic textbooks.

7. Studies similar to the present one with different combinations of factors, such as mental age and ability in fundamental operations, controlled.

REFERENCES

1. W. W. Charters, *Teaching the Common Branches*, pp. 292–93. Boston: Houghton Mifflin Co., 1913.

2. Leo J. Brueckner, "Diagnosis in Arithmetic", *Educational Diagnosis*. The Thirty-fourth Yearbook of the National Society for the Study of Education. Bloomington, Ill.: Public School Publishing Co., 1935.

3. Leo J. Brueckner, "The Social Phase of Arithmetic Instruction", Chapter VII, pp. 140–156. The National Council of Teachers of Mathematics, *Arithmetic in General Education*. New York: Bureau of Publications, Teachers College, Columbia University, 1941.

4. John Dewey, *How We Think*, Chapter I. Boston: D. C. Heath and Co., 1910.

5. Harry G. Wheat, *The Psychology of the Elementary School*, p. 153. New York: Silver Burdett and Co., 1931.

6. Progressive Education Association, Committee on the Function of Mathematics in General Education, *Mathematics in General Education*, pp. 75–76. New York: D. Appleton–Century Co., 1940.

7. Unpublished study by the writer.

8. See "Arithmetic" in *Review of Educational Research*, October, 1931; February, 1935; and December, 1937. Also, *A Critical Summary of Research Relating to the Teaching of Arithmetic*, by Walter S. Monroe and Max D. Engelhart. (See below)
9. Walter S. Monroe and Max D. Engelhart, *A Critical Summary of Research Relating to the Teaching of Arithmetic*, Bulletin No. 58. Bureau of Educational Research, College of Education. Urbana, Ill.: University of Illinois, 1931. pp. 115.
10. *Ibid.*, p. 96.
11. *Ibid.*, p. 97.
12. G. M. Wilson, "Arithmetic", *Encyclopedia of Educational Research*, p. 54. 1941.
13. Leo J. Brueckner et al. Published by the Educational Test Bureau, Minneapolis.
14. Published by the Public School Publishing Co., Bloomington, Ill.
15. Published by the World Book Company.
16. Published by the Bureau of Publications, Teachers College, Columbia University, New York.
17. Published by the Educational Test Bureau, Minneapolis, Minn.
18. Using the formula, $T = 10\frac{(x - \bar{x})}{S.D.} + 50$.
19. R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural, and Medical Research*. London: Oliver and Boyd, 1938.
20. C. J. Hoyt, "Note on a Simplified Method of Computing Test Reliability", *Educational and Psychological Measurement*, 1: 93–95, January, 1941.
21. George W. Snedecor, *Statistical Methods*, p. 183. Ames, Iowa; The Collegiate Press, Inc., 1938.
22. R. A. Fisher, *Statistical Methods for Research Workers*. Edinburgh, Scotland. Oliver and Boyd, 1938.
23. *Ibid.*
24. Palmer O. Johnson and J. Neyman, "Tests of Certain Linear Hypotheses and Their Application to Some Educational Problems", *Statistical Research Memoirs*, 1: 72–93, June, 1936.

THE EFFECT OF INSTRUCTION IN MATHEMATICAL VOCABULARY UPON PROBLEM SOLVING IN ARITHMETIC[1]

HARRY C. JOHNSON
State Teachers College at Buffalo

Editor's Note: Much progress has been made in ferreting out and making known difficulties in learning arithmetic. The author adds to this literature by indicating the importance of vocabulary in arithmetical achievement.

INTRODUCTION

DIAGNOSTIC studies in the field of problem solving in arithmetic have revealed numerous possible causes of difficulty in this important area of learning. Many experimental studies have already been made to determine the values of certain types of remedial measures toward the improvement of the problem solving ability among elementary school children.

The experiment summarized in this article limits itself to one narrow phase of the general difficulty that problem solving presents. Specifically, the question to be considered is, "What effect does the learning of mathematical vocabulary have upon the solution of numerical problems in which these words play an important part?"

Whether or not there is a relation between the knowledge of general vocabulary and problem solving ability in arithmetic was not the concern of this present experiment. Evidence would not be hard to obtain, to show that there is a considerable relationship between knowledge of specific mathematical terms and the solution of problems involving those terms. For example, a child would not be apt to solve correctly a problem calling for the area of a rectangle, unless he knew the meaning of "area", and what process to use in finding it. Similarly, problems using words such as discount, net proceeds, profit, ratio or average could not be attacked intelligently unless the meanings of these words were well understood.

It is, however, the purpose of this study to determine whether *improvement* in specific mathematical vocabulary leads to an *improvement* in the solution of problems which involve the use of these specific mathematical terms.

[1] Abstract of a doctor's dissertation written under the direction of Dr. Leo J. Brueckner, University of Minnesota.

To be able to demonstrate statistically the apparently obvious fact, that instruction in vocabulary leads to an improvement in the understanding of the meanings of these words, is certainly not to be minimized. To be able to show, further, that such learning is also accompanied by improvement in the solution of problems based upon these words is of much greater significance. Such evidence should tend to lay greater emphasis on the acquisition of specific vocabulary in mathematics, particularly, and possibly in other fields where technical vocabularies exist.

PROCEDURE

Homogeneity of the Schools.—This experiment, conducted during 1941–42 in three junior high schools of Niagara Falls, New York, involved 898 pupils in 28 seventh grade classes. These children were taught by a group of 15 teachers, with varying degrees of training and experience.

In order to justify the combining of the populations from the three schools into one large group, it was necessary to demonstrate that these schools were statistically homogeneous. The following tests were used: the Welch–Nayer test[2 and 3] for the homogeneity of variability, and the analysis of variance test[4] for the homogeneity of means. It was found that the schools were homogeneous in these two respects, on each of the following measures:

1. Chronological age.
2. Mental Age. (Henmon–Nelson Intelligence Test).
3. Reading ability. (Gates Reading Test–Ability to Read to Note Details).
4. Arithmetic problem solving and arithmetic vocabulary (Analytical Scales of Attainment).
5. Knowledge of arithmetic vocabulary studied prior to the beginning of the experiment.[5]

After it was demonstrated that the three schools came from the same general population, the various classes were assigned to the experimental

[2] B. L. Welch, "Note on an Extension of the L_1 Test," *Statistical Research Memoirs*, I (June, 1936), 52–56.

[3] P. P. N. Nayer, "An Investigation into the Applications of Neyman and Pearson's L_1 Test, with Percentage Limits," *Statistical Research Memoirs*, I (June, 1936), 38–52.

[4] E. F. Lindquist, *Statistical Analysis in Educational Research*. New York: Houghton Mifflin Company, pp. 62–65.

[5] Based upon results from a 40 point objective test of the multiple choice type, constructed by the writer.

or control groups. Ordinarily, this should be done in a random manner. However, since ability grouping was practiced in the school system employed in this experiment, the segregation of the classes into the two groups was accomplished in each school by alternating the classes from high ability to low ability.

The statistical procedure, chosen by the writer for the analysis of the data, did not require an individual-for-individual matching of the pupils in the experimental and control groups. Recently, techniques[6] have been developed which make it possible to adjust for mean differences that exist between groups with respect to basic matching characters. If however, it can be shown that the experimental and control groups are equivalent at the beginning of the experiment, the analysis of the data can be carried out quite satisfactorily through the application of less rigorous statistical techniques.

Equivalence of the Experimental and Control Groups.—A very comprehensive study was made of the experimental and control groups to determine whether or not they were comparable at the beginning of the experiment. The writer felt that the most important criteria to be considered were those knowledges and skills in the field of arithmetic which were most closely associated with the specific problem to be investigated. The test chosen for this purpose was the Analytical Scales of Attainment,[7] which provides measures of ability in four phases of arithmetic:

1. Understanding of quantitative relationships.
2. Ability to solve arithmetical problems.
3. Understanding of mathematical vocabulary.
4. Ability in the fundamental operations.

All cases for whom complete data were not available on these tests were discarded from the study. This left 316 pupils in the experimental group and 282 pupils in the control group. The comparison of the means of these two groups revealed that there was no significant difference (at the 1% level) between them with respect to any of the following factors:

1. Chronological age.
2. Mental age. (Henmon–Nelson Intelligence Test).

[6] Palmer O. Johnson and J. Neyman, "Tests of Certain Linear Hypotheses and Their Applications to Some Educational Problems," *Statistical Research Memoirs*, I (June, 1936), 72–93.

[7] This test, developed by Martha Kellogg, L. J. Brueckner and M. J. Van Wagenen, is published by the Educational Test Bureau, Minneapolis, Minn.

3. Four abilities measured by the Analytical Scales of Attainment.
4. Reading ability. (Gates Reading Test—Ability to Read to Note Details).
5. Knowledge of arithmetic vocabulary studied prior to the beginning of the experiment.

With only one exception[8] the two groups were also found to be homogeneous with respect to variability on the above mentioned traits.

A further study of the experimental and control groups, made with respect to nationality and occupation of father, revealed that on these factors as well, the two groups were very much alike.

The Testing and Instructional Program.—One of the chief criticisms that the writer feels can be directed against many experimental studies is that they are too far removed from the actual work done in the classroom at the time the experiments are conducted. It was decided at the outset of this experiment that it would be integrated, as far as possible, with the regular work taught during the mathematics period. The children would thus come to recognize the testing and remedial program as a very definite part of their own class work. It was felt that this approach would provoke greater interest in the experimental procedure, and would lead to a greater degree of effort and cooperation on the part of the children.

At a preliminary meeting with the teachers and the supervisor of mathematics, it was agreed that the experiment would extend over a period of about 14 weeks, and that it would confine itself to the content covered by five chapters in the text-book[9] which was uniformly used in all of the classes. These chapters dealt with the following general topics: decimal fractions, percentage, measurement, graphs and common uses of percentage. It was decided that each teacher would follow the same order of topics, and that these topics would be taught to the various classes, as far as possible, at about the same time.

The experiment was divided into three periods, each of which was preceded and followed by tests in vocabulary and problem solving, based upon the material taught during the respective periods of the experiment. The following are sample questions taken from the tests used:

[8] The variability in mental age of the experimental group was slightly higher than that of the control group.

[9] Betz, William, *Junior High School Mathematics*, Book One. New York: Ginn and Company, 1933.

Vocabulary Test: ***Perimeter*** means: (a) the amount of surface, (b) volume, (c) distance around, (d) distance across.
Problem Test: Find the *perimeter* of the rectangle whose length is 12 inches and whose width is 8 inches.

The tests were so constructed that for each word tested in the vocabulary test, there was a problem in the problem test which made use of and required the understanding of that word. The above test items concern themselves with the word "perimeter". Among other words, selected from the textbook material and incorporated in the test questions, were the following: receipts, balance, average, portion, area, fraction, income, successive, dimensions, diagonal, ratio, premium, marked down, mill and quart. In all, 60 items of vocabulary were used in the three sets of tests.

It should be pointed out that the problems were constructed in such a manner that there was a minimum emphasis on the computational skills involved. Once the child understood the meaning of the critical word, and the process of computation which it suggested, the solution of the problem would present no unusual difficulty. The scoring of the tests allowed no partial credit. Each problem was given a value of one point for the correct answer.

It was felt that some value might be derived by investigating the psychological principal known as the transfer of training. Let it be assumed that the remedial exercises which were designed to bring about improvement in the understanding of words found in the text-book actually accomplished that purpose. Would there, then, be any transfer of learning to words from the text which were not so emphasized and clarified by the remedial exercises? To study this phase of the problem, ten vocabulary items were added to each of the vocabulary tests, and were designated "transfer items".

The coefficients of reliability for the tests referred to above, computed by the split-halves method and corrected by the Spearman–Brown formula, varied from .77 to .91. According to accepted practice in educational research, these reliabilities are high enough for group measurements, such as those involved in this experiment.

The Instructional Materials for Experimental Classes.—The purpose of the practice exercises, designed for use in the experimental classes, was to develop a meaningful understanding of vocabulary beyond that which was provided by the text-book itself. The following types of exercises were considered:

1. Daily, oral drills on selected words.
2. Individuals use of dictionary to obtain word meanings.
3. Use of individual notebooks, in which pupils would record meanings of difficult words.
4. Mimeographed instructional materials and drills to supplement the text-book.

Whatever the treatment decided upon, it would be essential, for the sake of unity in the experiment, to prescribe the same type, or combination of types, to all of the experimental classes. Furthermore, should it be desirable to make comparisons between classes, it would also be essential for the same vocabulary to be emphasized in all classes, and as far as possible, to the same degree.

In the light of these criteria, it seemed that mimeographed materials, prepared by the writer, would provide the most desirable type of remedial treatment. These exercises were so designed that words relating to a common topic were grouped and discussed so as to bring out their individual meanings and their inter-relationships. Exercises, principally of the recall and matching types, accompanied each set of words, and were designed to give a reasonable amount of drill in the use of the words. Below is a sample of the type of exercise provided:

Area refers to *surface,* and is measured in *square units,* such as square inches, square yards, etc.

Perimeter means *distance around,* and refers to *length.* It is measured in inches, feet, miles, etc.

Volume refers to the amount of *space,* or *contents,* or *capacity.* It is expressed in terms of cubic inches, cubic feet, etc.

Exercises: Opposite each of the following, write *volume, area,* or *perimeter,* as the case may be.

(a) 350 square feet _____ (c) 3 square miles _____
(b) 40 yards _____ (d) 35 cubic inches _____

Tell whether each of the following refers to *volume, area,* or *perimeter:*

(a) amount of molding for a picture frame _____
(b) amount of wall paper for a room _____
(c) length of curb stone around a city block _____
(d) amount of water in an acquarium _____
(e) size of a state or county _____
(f) truck load of sand _____

A different set of these exercises was constructed for each period of the experiment. Each child in every experimental class was provided with his own copy. The time devoted to this instruction, averaging from five to eight minutes per class period, was taken from the regular class period; no extra time was allowed. The control classes had to rely entirely upon the text-book and regular class discussion for their learning of these mathematical terms.

THE FINDINGS

Final Status With Respect to General Learnings in Arithmetic.—It has already been pointed out that at the beginning of the experiment, the experimental and control groups were equivalent with respect to all of the abilities measured by the Analytical Scales of Attainment. A comparable form of this test was administered at the close of the experiment, to obtain measures of growth in these general areas. A study of these two sets of data revealed that significant gains were achieved during the course of the experiment by both the experimental and control groups, in all four of the areas of arithmetic included in this test. With respect to these gains, however, there were no significant differences in favor of either the experimental or the control group.

Final Status With Respect to Specific Learnings in Arithmetic.—By specific learnings, reference is made to the learnings measured by the vocabulary and problem tests prepared by the writer. The first comparison to be made pertains to the scores on the three vocabulary tests, with particular reference to those words which were included in the remedial exercises used by the experimental classes. Table I, tabulating these data, indicates that on the first pre-test the mean scores made by the experimental and control groups were 12.39 and 12.17, respectively. These scores reveal a difference of .22 of a point in favor of the experimental group. In order for this difference to be statistically significant, at the one per cent level, the ratio of the difference to its standard error (represented in the table by the letter k) would have to be 2.58. Since the value of k in this instance was only 1.02, it is obvious that the difference between the two groups on pre-test I was not real difference, and that it might well have arisen through errors of random sampling.

Similarly, it is seen from a study of Table I that the experimental and control groups were statistically equivalent on the basis of the tests administered before each of the remaining two periods of the experiment.

TABLE I

COMPARISON OF INITIAL AND FINAL SCORES OF EXPERIMENTAL AND CONTROL GROUPS WITH RESPECT TO KNOWLEDGE OF VOCABULARY TAUGHT IN REMEDIAL EXERCISES DURING EACH PERIOD OF THE EXPERIMENT

(Total Score on Each Test = 20 Points)

Test	Group	n	$\bar{x}$	S. D.	S.E. $_{\bar{x}}$	Comparison of Means		
						Diff. $\bar{e} - \bar{c}$	S. E. $_{diff.}$	k
I Pre	Exper.	307	12.39	2.59	.147			
	Control	271	12.17	2.59	.158	+ .22	.215	1.02
I Post	Exper.	307	16.14	2.46	.140			
	Control	271	13.75	2.76	.168	+2.39	.219	10.90
II Pre	Exper.	280	11.40	2.90	.173			
	Control	247	11.79	2.70	.172	— .39	.244	1.60
II Post	Exper.	280	16.00	2.32	.139			
	Control	247	13.81	2.62	.167	+2.19	.217	10.58
III Pre	Exper.	280	12.14	2.59	.155			
	Control	245	11.86	2.87	.183	+ .28	.240	1.17
III Post	Exper.	280	15.99	2.87	.171			
	Control	245	14.93	2.59	.165	+1.06	.237	4.48

On each of the post-tests, however, substantial differences, each of which is statistically significant, are seen to be in favor of the experimental group. The greatest gains in favor of the experimental group were made during the first and second periods of the experiment.

Two reasons might be cited to account for the less pronounced gain in favor of the experimental group during the last period of the experiment. In the first place, circumstances beyond the control of the experimenter necessitated a reduction in the amount of time devoted to the remedial exercises during the final period of the experiment. Secondly, with respect to the vocabulary of the third period, dealing almost exclusively with common business uses of percentage, the text-book was quite descriptive and complete in giving meanings and understandings. The control group, therefore, had the benefit of much the same type of instruction that the experimental group received through the use of the remedial exercises.

Gains in Untaught Vocabulary.—Now that it has been demonstrated that the experimental group showed significantly greater growth than did the control group in the knowledge of those words included in the remedial

exercises, the question might well be asked, "Did similar differences arise with respect to other words from the text-book which were not included in the remedial exercises?" In other words, what evidence was there of a transfer of learning from taught words to untaught words? In Table II the means for both experimental and control groups are given on the transfer items for each period of the experiment. For the post-tests, as well as for the pre-tests, the differences between the means of the two groups are much too small to be statistically significant. It seems apparent, in the face of these facts, that in order to bring about improvement in the understanding of specific words, it is necessary to teach those words in a very specific manner. There is no evidence, at least from the results of this experiment, that a transfer of training exists in this area of learning.

TABLE II

COMPARISON OF INITIAL AND FINAL SCORES OF EXPERIMENTAL AND CONTROL GROUPS WITH RESPECT TO KNOWLEDGE OF VOCABULARY NOT TAUGHT IN REMEDIAL EXERCISES DURING EACH PERIOD OF THE EXPERIMENT

(Total Score on Each Test = 10 Points)

Test	Group	n	$\bar{x}$	S. D.	S.E.$_{\bar{x}}$	Comparison of Means		
						Diff. $\bar{e} - \bar{c}$	S. E.$_{diff.}$	k
I Pre	Exper.	307	5.90	1.53	.087			
	Control	271	5.74	1.57	.095	+ .16	.129	1.24
I Post	Exper.	307	6.52	1.47	.084			
	Control	271	6.35	1.45	.088	+ .17	.122	1.39
II Pre	Exper.	280	5.91	1.58	.094			
	Control	247	5.97	1.60	.102	— .06	.138	.43
II Post	Exper.	280	7.42	1.67	.100			
	Control	247	7.26	1.56	.100	+ .16	.141	1.14
III Pre	Exper.	280	5.73	1.39	.083			
	Control	245	5.56	1.50	.096	+ .17	.127	1.34
III Post	Exper.	280	7.38	1.74	.104			
	Control	245	7.48	1.61	.097	— .10	.142	.70

Gains in Problem Solving.—The growth of the two groups in the solution of problems based upon the taught vocabulary will now be considered. Table III shows that on each of the pre-tests, the differences between the groups were not large enough to be statistically significant at the one per cent level (all k values are less than 2.58). At the end of the first

and second periods, however, the differences in favor of the experimental group, even though they were not striking, were large enough to be statistically significant. These differences were .96 and 1.44 points, respectively, out of a total of 20 points on each of the tests. The actual gains, however, are somewhat less, when the initial scores of the two groups are taken into consideration.

TABLE III

COMPARISON OF INITIAL AND FINAL SCORES OF EXPERIMENTAL AND CONTROL GROUPS WITH RESPECT TO SOLUTION OF PROBLEMS INVOLVING VOCABULARY TAUGHT IN REMEDIAL EXERCISES DURING EACH PERIOD OF THE EXPERIMENT

(Total Score on Each Test = 20 Points)

Test	Group	n	$\bar{x}$	S. D.	S.E.$_{\bar{x}}$	Comparison of Means		
						Diff. $\bar{e} - \bar{c}$	S. E.$_{diff.}$	k
I Pre	Exper.	307	10.83	3.80	.188			
	Control	271	10.80	3.26	.198	+ .03	.273	.11
I Post	Exper.	307	12.86	3.60	.205			
	Control	271	11.90	3.29	.200	+ .96	.286	3.36
II Pre	Exper.	280	6.51	3.02	.181			
	Control	247	5.98	2.53	.161	+ .53	.242	2.19
II Post	Exper.	280	9.62	3.49	.208			
	Control	247	8.18	2.94	.187	+1.44	.280	5.14
III Pre	Exper.	280	10.32	4.23	.253			
	Control	245	10.66	3.99	.255	— .34	.359	.95
III Post	Exper.	280	14.51	3.85	.230			
	Control	245	14.50	3.65	.233	+ .01	.327	.03

The third period of the experiment did not yield results comparable to those of the first and second periods. Even though the net gain of the experimental group beyond that of the control group was .35 of a point (from — .34 to + .01), this gain was not significant. This situation is somewhat analogous to that which was observed earlier, when the gains on the vocabulary tests were discussed. It is highly possible that the explanation given at that time for the less pronounced difference between the two groups, during the third period, might also be applicable to the present situation.

Determination of Regions of Significance.—Were the gains that were discovered in favor of the experimental group common to all levels of

ability, or were they limited to pupils of certain levels of ability only? It is highly important to answer this question, particularly because of its implications in the field of remedial teaching.

One method of obtaining the answer to this question would be to stratify the samples, and compare means of the two groups at a series of different levels. This would be a lengthy task, and would not under ordinary conditions yield highly satisfactory results, particularly if the sizes of the samples were small.

A technique developed by Johnson and Neyman[10] can be used to solve this problem without stratifying the samples. The application of this technique enables the experimenter to determine, first of all, whether or not significant differences exist between two groups when they are matched statistically on two basic characters, and secondly, it makes it possible to locate that region where-in the differences are significant.

In making this analysis, the data from the three periods of the experiment were combined. One investigation compared the control and experimental groups with respect to their total gains in taught vocabulary when the two groups were matched statistically on mental age and initial status in vocabulary. The other investigation concerned itself with a comparison of the two groups with respect to their total gains in problem solving when they were matched statistically on mental age and initial status in problem solving.

The writer will not go into the details necessary to present the complete solutions to these two statistical problems.[11] The findings indicated, however, that the experimental group achieved significantly greater gains than did the control group in both vocabulary and problem solving, and that this superiority of the experimental group maintained for pupils of practically all levels of mental ability and initial status in the area under study.

Other Findings.—Several other problems, inherent in or arising out of the major study, were also investigated. They are summarized briefly as follows:

[10] Op. cit.

[11] For the complete solution of a similar problem, the reader is referred to Robert H. Koenker and Carl W. Hansen, "Steps for the Application of the Johnson–Neyman Technique—a Sample Analysis," *Journal of Experimental Education*, X (March 1942) 164–73.

1. A rotation study, involving four classes and two teachers, carried on concurrently with the major experiment, tended to show that the use of remedial exercises in vocabulary, of the type used in this experiment, is likely to lead to increased learnings in vocabulary in whatever class they are used.

2. A study of gains in vocabulary and problem solving made by individual classes revealed that in practically every instance where experimental and control classes were taught by the same teacher, the greater gains were achieved by the experimental classes.

3. The results from the retention test administered three months after the close of the experiment revealed the fact that the experimental group was still statistically superior to the control group in the knowledge of the vocabulary included in the remedial exercises. During this interval, however, the difference between the two groups was narrowed slightly. This change was brought about partly by a loss experienced by the experimental group, and partly by a small gain realized by the control group.

4. An analysis of pupil and teacher reactions to the remedial exercises indicated that in the vast majority of cases, both pupils and teachers were highly in favor of the use of these materials. Furthermore, the teachers and also the pupils were strongly convinced that these materials were of considerable value in contributing to the educational growth of the pupils.

5. The efficacy of the remedial exercises employed in this experiment did not appear to be conditioned by the skill and personality of the teacher under whose guidance and direction they were used. Favorable results were achieved by teachers who were either limited in their experience or who had very little previous knowledge of the needs of the children they were teaching.

SUMMARY AND CONCLUSIONS

The fundamental purpose of the experiment described in the preceding paragraphs was to determine what effect specific instruction in vocabulary had on success in problem solving in seventh grade mathematics. Incidental to this purpose, yet almost as pertinent in value, was determining the effect this instruction in vocabulary had on the actual learning of the vocabulary itself. The following conclusions are suggested by the findings:

1. The use of instructional materials in mathematical vocabulary leads to significant growth in the knowledge of the specific terms included in

these materials, as well as in the solution of numerical problems involving the use of these terms. Values derived from such study are applicable to pupils of practically all levels of ability.

2. The use of vocabulary exercises of a specific nature, such as those employed in this experiment, does not tend to bring about a general improvement in arithmetical learnings.

3. There was no evidence to indicate that there is a transfer of training from the learning of words from the text-book which are taught through the use of practice exercises, to the learning of other words from the text-book which are not so taught and emphasized.

4. The value of instructional materials such as those used in this experiment is independent of the class in which they are used. On the whole, a teacher using such materials can bring about greater growth in vocabulary and problem solving than if she did not use them.

5. The successful use of remedial exercises of the type used in this experiment does not necessarily require the guidance and supervision of a well-experienced teacher. Beneficial results can be obtained when such materials are handled by teachers with rather limited experience.

6. In order to insure an adequate retention of learnings contributed by vocabulary instructional materials, it is necessary that they be used regularly and systematically as an integral part of the classroom procedure.

EDUCATIONAL IMPLICATIONS

The educational value of a program of study such as that organized and carried out in this experiment should be readily apparent to any teacher. That the values to the pupils were practical in nature and fairly permanent in their effect, was borne out by the findings of the experiment. Greater effects would no doubt be observed, extending even into the general areas of arithmetic, if such training in vocabulary were made an integral part of the work in arithmetic from the time when a formal study of the subject is first begun.

The writer has shown that the materials he used to bring about greater growth in vocabulary, and concomitantly in problem solving, were equally valuable to pupils of all levels of ability. Likewise, they seemed to be equally useful in the hands of both experienced and inexperienced teachers. These facts should recommend a general use of similar vocabulary-building exercises to all teachers of arithmetic.

An examination of any recent series of text-books in elementary arithmetic will reveal its author's concern over the difficulties that children have in problem solving. Very little emphasis, however, has been placed on the adequate building of the vocabulary of arithmetic. Only rarely will one find an exercise whose purpose is to test the understanding of the crucial words presented in the reading matter. Authors of text-books in arithmetic, or any other branch of mathematics, would do well to apply the findings of this experiment to future revisions of their publications. Work-books, designed for the sole purpose of providing tests and practice materials in the vocabulary of arithmetic, should find a ready market.

PROBLEMS SUGGESTED FOR FURTHER STUDY

The following problems, arising out of the present study, might prove profitable as subjects for further research:

1. Would results similar to those obtained in this study be found if vocabulary training were made an integral part of the teaching of arithmetic in other grades of the school?

2. How far-reaching would be the benefits of vocabulary exercises if their use were extended over a period of several years?

3. What values might be derived from other types of practice exercises designed to build an understanding of vocabulary, such as the use of the dictionary, the recording of meanings of important words in pupil notebooks, and other forms of word study?

4. To what extent would the study of vocabulary in science, geography, history and other subjects, aid the pupil in meeting problem situations that arise in those areas of learning?

FACTORS ASSOCIATED WITH SUCCESSFUL ACHIEVEMENT IN PROBLEM SOLVING IN SIXTH GRADE ARITHMETIC

CARL W. HANSEN
Quincy Public Schools, Quincy, Illinois

Editor's Note: The last twenty or twenty-five years have seen great progress in our understanding of the factors conditioning the learning of the school subjects. The author contributes to our understanding of these factors as they relate to the learning and teaching of arithmetic.

INTRODUCTION

THIS discussion concerns itself primarily with an analysis of certain skills and abilities of pupils in relation to their ability to solve verbal arithmetic problems of the kind included in standardized tests. A survey of recent educational literature suggests that some relationship exists between these skills and success in solving verbal arithmetic problems. Despite the seeming indication of this relationship, close examination of the literature points to the fact that in most instances the evidence is inconclusive. Opportunities for research in this field, are, therefore, numerous. That research is urgently needed is generally conceded. In fact, continuous research is essential if advancement is to be made in such areas of arithmetic instruction as educational diagnosis, remedial instruction, and curriculum revision.

With this situation in mind the writer[1] endeavored to determine whether certain abilities are associated with superior achievement in solving verbal arithmetic problems. Tests[2] in problem solving were administered to 681 sixth grade pupils in ten communities ranging in size from an agricultural village of 400 inhabitants to an industrial and commercial city of 25,000. After the frequency distributions for the ten communities were found to be homogeneous with respect to variability and mean achievement, they were grouped in one composite distribution which included the entire population. The upper 27 per cent, the 184 best problem solvers, were designated superior achievers and the lower 27 per cent, the 184 poorest problem solvers, were designated inferior achievers.

---

[1] Hansen, Carl W. *Factors Associated with Superior and Inferior Achievement in Problem Solving in Sixth Grade Arithmetic: A Statistical Analysis.* Unpublished Doctor's Dissertation, University of Minnesota, June 1943.

[2] a. Problems Test, Analytical Scales of Attainment; Arithmetic, Division 2, Form 2.

b. Test 14, Problems, California Test of Mental Maturity, Elementary Series.

Since it was obviously impossible to consider all the factors suggested as being related to successful achievement in verbal problem solving in arithmetic, the two groups, superior and inferior achievers, were compared statistically on selected factors thought to be related to success in arithmetic problem solving, classified arbitrarily as follows:

Arithmetical Factors	*Test Used*
Fundamental Operations	Analytical Scales[3]
Quantitative Relationships (social)	Analytical Scales
Arithmetic Vocabulary	Analytical Scales
Estimating Answers to Problems	National Achievement[4]
Estimating Answers to Examples	National Achievement
Problem Analysis	National Achievement
Finding the Key to a Problem	National Achievement
Thinking Abstractly With Numbers	California[5]
Number Series	California
Mental Factors	
General Reasoning Ability	California
Noting Differences	California
Noting Likenesses	California
Non-Language Factors	California
Analogies	California
Delayed Memory Span	California
Immediate Memory Span	California
Memory	California
Spacial Imagery	California
Spacial Relationships	California
Inference	California
Reading Factors	
General Vocabulary	California
General Language Ability	California
Speed-Reading to Predict Outcomes	Gates[6]
Comprehension-Reading to Predict Outcomes	Gates

[3] *Analytical Scales of Attainment-Arithmetic:* Division 2, Form A. Minneapolis, Minn. Educational Test Bureau.

[4] *National Achievement Tests*—Arithmetic Reasoning and Fundamentals (For Grades 3 to 8). Rockville Center, New York. Acorn Publishing Company.

[5] *California Test of Mental Maturity*—Elementary Series. Los Angeles: California Test Bureau.

[6] *Gates Silent Reading Tests,* Types B and D, Grades 3–8, Form 2. New York: Bureau of Publications, Teachers College, Columbia University.

Reading Factors	*Test Used*
Speed-Rating to Note Details	Gates
Comprehension-Reading to Note Details	Gates
Reading Graphs, Charts and Tables	Iowa[7]

The two groups were first compared on the above factors by means of the "t" test for the significance of the difference between two means. This comparison showed that the superior group was significantly better than the inferior group on all the factors listed. Since, however, the successful achievers might be superior to the inferior group because of such variables as mental age and chronological age, it was necessary to compare the two groups when the effects of these variables were ruled out. This can be done by a "matching technique". However, this approach is of limited value since so many cases are lost in matching that the final groups are usually no longer representative of the original groups. A more satisfactory procedure is the application of the Johnson–Neyman technique.[8] By use of this technique it is possible to test the statistical significance of the best estimate of the difference in achievement between the two groups, when the two groups are matched statistically on two basic characters. Application[9] of the technique, therefore, eliminates the need of an "individual-for-individual" matching procedure. In addition, it has a further advantage in that it is also possible to determine the limits of the regions wherein the differences between the groups are significant. The two groups, superior and inferior achievers, were therefore matched statistically, the basic matching criteria, mental age and chronological age being controlled by application of the Johnson–Neyman technique.

FACTORS IN WHICH SUPERIOR ACHIEVERS EXCEL

When the two groups were matched with the basic matching criteria under statistical control, the superior group excelled in all except four factors, these measured by the Gates Reading Tests B and D, namely, speed in

[7] 1938 Iowa Every-Pupil Test of Basic Skills, Grades 6, 7, and 8, Test B-Vocabulary: Basic Study Skills. Bureau of Educational Research and Service, State University of Iowa, Iowa City, Iowa.

[8] Palmer O. Johnson and J. Neyman, "Tests of Certain Linear Hypotheses and Their Application to Some Educational Problems". *Statistical Research Memoirs*, I. June 1936, 72–93.

[9] Robert H. Koenkerr and Carl W. Hansen, "Steps for the Application of the Johnson–Neyman Technique—A Sample Analysis." *Journal of Experimental Education*, X, March 1942, 164–173.

reading to predict outcomes, comprehension in reading to predict outcomes, speed in reading to note details, and comprehension in reading to note details. With two exceptions the level of significance for the adjusted difference was in favor of the superior achievers at the one per cent level. Stated differently, this means that there is less than one chance in one hundred that a difference as large as the adjusted difference could have been due to errors in sampling.

TABLE I

LEVEL OF SIGNIFICANCE AND ADJUSTED DIFFERENCE BETWEEN THE MEANS OF THE SUPERIOR AND INFERIOR ACHIEVERS WITH RESPECT TO THE FACTORS SELECTED FOR STUDY WHEN THE BASIC MATCHING CRITERIA HAVE BEEN STATISTICALLY CONTROLLED

Factor	Level of Significance (Per Cent)	Adjusted* Difference
Arithmetical Factors		
Fundamental Operations	1	4.11
Estimating Answers to Examples	1	1.31
Ability to Solve Problems	1	41.78
Thinking Abstractly with Numbers	1	2.58
Estimating Answers to Problems	1	1.70
Problem Analysis	1	3.45
Number Series	1	.92
Quantitative Relationships (social)	1	3.49
Finding Key to a Problem	1	.91
Arithmetic Vocabulary	1	.99
Mental Factors		
General Reasoning Ability	1	7.42
Noting Differences	1	1.97
Noting Likenesses	1	.78
Non-Language Factors	1	9.20
Analogies	1	.92
Delayed Memory Span	1	1.51
Immediate Memory Span	1	1.02
Memory	1	2.52
Spacial Imagery	1	.62
Spacial Relationships	5	1.28
Inference	1	.85
Reading Factors		
General Language Ability	1	2.68
Reading Graphs, Charts & Tables	1	.15
General Vocabulary	5	.30
Speed-Reading to Predict Outcomes	1	—.88#
Comprehension-Reading to Predict Outcomes	----	—.45#
Speed-Reading to Note Details	----	—.88#
Comprehension-Reading to Note Details	----	—.54#

*By adjusted difference is meant the difference between the means of the two groups when the effect of the basic matching criteria, Mental and Chronological Age is eliminated statistically. Differences are based on scale values and are therefore not comparable from one factor to another.

Favors inferior achievers.

Arrangement of the factors in Table I, while arbitrary, serves to emphasize the character of the items which are apparently most closely related to success in solving verbal arithmetic problems. While it is not definitely

known what constitutes the ability to solve verbal arithmetic problems, it is fairly evident that those factors for which a significant adjusted difference exists are undoubtedly important elements in that ability. In general, the factors most closely associated with superior achievement appear to be those classified under the headings of (1) arithmetical factors, and (2) mental factors. A significant adjusted difference at the one per cent level in favor of the superior achievers was found for two items grouped under "reading factors", namely, general language ability, and the reading of graphs, charts, and tables. The difference in general vocabulary was significant at the five per cent level. A significant adjusted difference in favor of the *inferior* achievers was found for speed in reading to predict outcomes. No significant difference was found in favor of either group for the remaining three reading items.

This seeming lack of relationship between certain reading abilities and successful achievement in problem solving may lead to the conclusion that skill in general reading and knowledge of general vocabulary are not essential for success in verbal problem solving in arithmetic, and that reading skills and vocabulary in arithmetic are specific in that field. This conclusion may be warranted, but two factors inherent in the original investigation must be considered, namely, (1) the predominantly verbal character of the test[10] used to determine mental age, and (2) the relatively few reading and vocabulary factors considered in relation to the whole field of reading. While these limitations necessitate withholding final judgment, indications are that beyond the mastery of fundamental reading skills and basic vocabulary it may be more profitable to devote instructional time allotted to the teaching of verbal problem solving in arithmetic to abilities bearing more directly on arithmetic and reasoning. It should be pointed out that the Gates tests do not include mathematical materials, but literary content. The relationship found here should be studied further by means of materials more definitely mathematical in nature.

Obviously success in solving arithmetic problems is more likely to occur when teaching methods designed to contribute to success are undertaken as a part of the instructional program. Evidence presented in Table I indicates that the abilities in which the superior achievers were significantly

[10] *Otis Quick-Scoring Mental Ability Tests, Beta Test: Form A.* The verbal character of this test may have resulted in a disproportionate correction for the effect of Mental Age on the achievement of the groups.

better may be useful in organizing an instructional program of this character. Increased emphasis should clearly be given to those listed abilities shown to be related to successful achievement in verbal problem solving in arithmetic. For example, an instructional program intended to develop through concrete experiences and well-directed practice an understanding of arithmetical processes, clear number concepts, skill in mathematical computation and reasoning, and basic reading skills involving the use of arithmetical materials may materially reduce the number of poor achievers in problem solving. Likewise, a diagnostic procedure involving the same or similar factors may be employed to locate causes of disability, and, if followed by a proper remedial program, will likely result in greater instructional efficiency in the teaching of verbal arithmetic problems.

REGIONS OF SIGNIFICANCE

The Johnson–Neyman technique enables the investigator to determine not only the specific traits for which a statistically significant difference between good and poor achievers exists, but also to define the limits of the population in terms of the basic matching criteria. For the sample studied, the actual minimum and maximum mental ages were 88 months (7-1/3 years) and 218 months (18-1/6 years) respectively, while the minimum and maximum chronological ages were 120 months (10 years) and 200 months (16-2/3 years) respectively. Theoretical mental and chronological age limits derived by statistical analysis were in the case of some factors far in excess of those for the actual sample. In order to avoid confusion in the mind of the reader, the ranges in Table II are restricted to the actual mental and chronological age limits for the sample population.

Presentation of this type of information in tabular form involves a definite limitation since theoretical limits derived by application of the Johnson–Neyman technique are either elliptical or hyperbolic in nature while presentation in tabular form requires approximations essentially quadrangular in character. The limits presented, therefore, are not exact, but show in general the mental and chronological age ranges within which pupils may be expected to do significantly better if they are superior achievers in solving verbal arithmetic problems than if they are poor achievers.

These data are educationally significant in that they show the age limits within which teachers may expect success in teaching those factors shown to

be significantly related to successful achievement in problem solving. For some of the factors the region of significance is very small, for instance, inferences. For some of the factors the data indicate that an instructional program designed to develop ability in problem solving may be undertaken in the lower primary grades (lower mental age limits 88 months) while in the case of other factors (lower mental age limits 120 months or more) instruction should, no doubt, be delayed until the later years of the elementary school. Similar chronological age limits can readily be noted in the table.

TABLE II

NATURE OF REGION OF SIGNIFICANCE AND APPROXIMATE MENTAL AND CHRONOLOGICAL AGE LIMITS (IN MONTHS) WITHIN WHICH THE SUPERIOR GROUP IS SIGNIFICANTLY BETTER IN RELATION TO EACH FACTOR STUDIED

Factor*	Nature of # Region of Significance	Approximate Mental Age Limits		Approximate Chronological Age Limits	
		Min.	Max.	Min.	Max.
Arithmetical Factors					
Fundamental Operations	E	88	218	120	163
Estimating Answers to Examples	H	123	218	120	170
Ability to Solve Problems	E	88	218	120	200
Thinking Abstractly with Numbers	E	88	218	120	200
Estimating Answers to Problems	E	88	218	120	191
Problem Analysis	E	88	218	120	170
Number Series	H	130	218	120	200
Quantitative Relationships (social)	E	107	218	120	200
Finding Key to a Problem	E	88	148	127	161
Arithmetic Vocabulary	E	167	200	127	137
Mental Factors					
General Reasoning Ability	E	88	218	120	200
Noting Differences	E	106	163	130	164
Noting Likenesses	E	88	145	120	190
Non-Language Factors	E	88	202	140	200
Analogies	E	121	151	136	152
Delayed Memory Span	H	88	169	136	200
Immediate Memory Span	E	88	121	150	168
Memory	H	88	157	136	200
Spacial Imagery	E	88	140	132	170
Spacial Relationships	E	88	124	144	200
Inference	E	142	146	137	141
Reading Factors					
General Language Ability	E	88	218	140	200
Reading Graphs, Charts and Tables	H	133	218	120	200
General Vocabulary	H	156	218	130	200

* All factors listed are statistically significant in favor of the superior achievers at the 1 per cent level of significance except for Spacial Relationships and General Vocabulary which are at the 5 per cent level.

E—Ellipse.
H—Hyperbola.

Closely related to the instructional program is the problem of curriculum revision. If an instructional program is developed which will stress those factors shown to be associated with successful achievement in verbal problem solving, it is evident that the curriculum must be revised accordingly. Undoubtedly, it will be necessary to include within the curriculum experiences not now generally used in connection with formal instruction in verbal problem solving in arithmetic.

Further study on these factors would appear to be desirable. Investigation is urgently needed on the relationship between various reading skills and success in solving verbal arithmetic problems. Experimentation in practical classroom situations appears necessary to reconcile the findings based on statistical analysis and actual performance of pupils following the use of instructional procedures based on the evidence contained in this study.

BASIC CONSIDERATIONS FOR PROFITABLE RESEARCH IN ARITHMETIC[1]

GUY M. WILSON
Boston University

Editor's Note: With so much to be done it is always a problem to know upon what one's efforts might best be centered. The author raises this question relative to research on the teaching of arithmetic.

IT IS possible to have research on a purely academic basis, and while it may not be valuable positively, it can have a certain amount of negative value. For example, it is possible to study the errors which pupils make in division of decimals at upper grade levels. It is interesting to note[2] that 70% of the errors which do occur in the division of decimals are "peculiar to decimal situations". It is interesting to note that practically every type of error which can occur, does occur. It is interesting to note that decimal errors, grades 6–9 inclusive, can be grouped into twenty-one classifications; that errors resulting from placing the point or shifting the point are high and reasonably constant throughout these grades; that zero errors increase at the ninth grade level; and that all told, results are so poor as to suggest that something very radical should be done.

What are the possible radical procedures? One person suggests better teaching of division of decimals in the grades. Another suggests re-teaching of decimals at different upper grade and high school levels. But few specialists in arithmetic are willing to admit that the results suggest that the time on division of decimals in the grades has been wasted, probably worse than wasted, since many children are by such teaching discouraged from going further in the field of mathematics.

A recent study of errors in division of decimals, bases conclusions on an examination with abstract numbers which suggests drill as the method of teaching. Is drill appropriate on an example such as .51 | 107.1, at the sixth grade level? There are many phases of school work where drill is not justified, where, in fact, it is the wrong method to use. What are the criteria for drill?

[1] Read before a recent meeting of the Elementary Mathematics Club, Boston University.

[2] F. E. Grossnickle: "Kinds of Errors in Division of Decimals and their Constancy," *Journal of Educational Research*, XXXVII, October 1943, 110–117.

Criteria for drill may be set up approximately as follows:

1. Drill is not appropriate unless the subject matter on which mastery is undertaken, is *very useful,*—so useful that it comes into actual use almost daily,—at least very, very often. Illustrations of useful subject matter justifying drill are beginning reading, writing legibly, spelling the basic words (the first 1000 or even 2000), addition of whole numbers. Is division of decimals a process used so often as to justify attempts at mastery at the sixth grade level? This point will be examined further.

2. It is not drill unless used *exactly as is.* A word must be spelled one way only. An addition combination has one answer only.

Work in decimals meets this one requirement for drill.

3. If it is drill for which an examination may later be set up, then *perfect mastery* is the only standard. Tests in division of decimals uniformly show low-grade mastery. Before concluding that mastery must be attained at any cost, the question of usefulness as raised under the first criterion above, must be answered.

4. The drill *load must be small enough* to permit perfect mastery. Here enters a question of educational statesmanship. How much of the total school time can be justly allotted to arithmetic? Certainly not more than the one-sixth of total school time indicated by various studies of a few years ago (Payne, Holmes, Ayres). Since those studies, time on arithmetic has tended to decrease, but scores in essential processes have improved. In other words we have in recent years applied the criterion of usefullness more rigidly, and as a result have decreased the drill load. As a result children have been securing better scores.

The four criteria indicated above are sufficient, if properly applied, to greatly improve the undertaking in arithmetic. But there are other requirements,—the requirements of good teaching. Such requirements would insist that drill should not begin too soon, and that it should never be undertaken without full and adequate provision for meaning and motivation. This is a full story in itself and will not be undertaken here.

To return to decimals, how useful is the process of division of decimals to the ordinary citizen in daily life? Fortunately there are adequate studies on the usefulness of decimals, and the circumstances under which the decimal processes are used.

Dalrymple[8] made a careful study of the uses of decimals in industry and in periodicals. In such industries as the American Sugar Refining Com-

[8] Marion Dalrymple: *A Study of the Present Uses of Decimals in Industries, Periodicals, and Textbooks.* Master Thesis, Boston University, School of Education, 1933.

pany and the Gillette Safety Razor Company she found that a large majority of the workers had no use for decimals, that some needed a reading knowledge of decimals, and that a few figured in decimals, but that this figuring was highly specialized and was done by highly trained people. It is improbable that anything could have carried over to this job from work on decimals in the grade schools. Dalrymple recommends that any teaching of decimals in the schools be limited to general understanding and a reading knowledge. Her study of decimals in periodicals lead to the same conclusion.

Buckley[4] made an intensive study of the uses of decimals in business. In general trade and barter, he found no decimals other than United States Money. So he went to highly specialized businesses. In his study in the Boston area, he included Chandler & Company, Forest Hills Hospital, Otis Elevator Company, Devonshire Financial Service Corporation, Burroughs Adding Machine Company, United Drug Inc., and Whiting Milk Company.

Even in these highly specialized businesses, Buckley found that 76.51% of all employees had no uses for decimals. There were 2.36% who needed a reading knowledge of decimals. Those who computed in decimals, 21.13% of the total, were highly trained, many with four years of university work in engineering, accounting, or hospital training. The computations were very specialized and for the most part were learned on the respective jobs. Buckley concludes that "a broad reading knowledge of decimals is more necessary than computational knowledge. Most of the decimal figuring done in specialized businesses is decimalization of percentages, mostly three place. A few, about 2.5%, need a highly technical or expert knowledge of decimals."

Russell's[5] study of decimals is extensive enough to be authoritative. Aside from United States money, Russell found practically no use for decimals in common usage. So he went to big business and he covered the twenty-three main census classifications in a study that involved eighty-nine industrial plants and over sixty-eight thousand employees. Of these sixty-eight thousand employees, 88.2% had no uses for decimals, 4.6% needed a reading knowledge, and 7.4% needed to figure in decimals. Most figuring is highly special-

[4] J. E. Buckley: *Uses of Decimals in Business*. Master Thesis, Boston University, School of Education, 1935.

[5] Garland Russell: *Decimal Usage in the Occupational World*. Doctor's Dissertation, Boston University, 1943.

ized; it has been learned on the job or acquired through special training. There would be little or no carry over from work on decimals in the grades.

Russell recommends that the work on decimals in the grades or general high school be confined to general understanding and a reading knowledge.

These studies by Dalrymple, Buckley, and Russell indicate clearly that division of decimals is not a proper drill assignment for grade or junior high school pupils as a part of general mathematics. Decimals in any form (aside from United States money) are seldom used by the ordinary citizen. Decimal figuring is very special figuring. It has been learned on the job or at the college and university level. The engineering student at the university level may need decimals, but he carries over nothing from his grade experience with drill on decimals, which could not have been secured more easily and more meaningfully without drill.

The better method for a seldom used process is appreciation. Harap[*] has shown an appreciation unit in decimals which carried meaning to sixth grade pupils without requiring drill mastery. An appreciation unit on decimals could be centered around the history of decimals. The unit fraction was practically the only fraction of the ancient and mediaevel world; the common fraction, with numerators other than one, finally made a place for itself. Credit for the discovery of decimal fractions goes to Simon Stevin of Belgium, 1584. Three decades later, Napier gave us the decimal point. Thus it was one thousand years after the Hindu decimal system of number was introduced into Western Europe, before it was finally extended to the right to include decimal fractions. Here is the basis for a fine appreciation unit on decimals, without any mastery requirement, but with good opportunities for understanding. The addition, subtraction, multiplication and divisions of decimals as usually taught in the intermediate or upper grades has little or no value. It is too frequently drill without meaning. It consumes too much time, and the method is not well chosen. This same work developed as a short appreciation unit with emphasis upon insight and meaning, would at least have the merit of doing little harm; if taught only to the brighter pupils, or "those especially interested in arithmetic", it could have real values.

---

[*] Henry Harap and C. E. Mapes: "Learning of Decimals in an Arithmetic Activity Program", *Journal of Educational Research*, XXIX (May, 1936), 686–693.

The only requirements for decimals for grade pupils should be, general understanding and a reading knowledge. There should be no drill requirements, and no drill examination.

Returning now to the question of research in arithmetic, it is apparent:

1. That not all arithmetic is drill.

2. That the remedy for poor drill results may not always be more drill. Sometimes the abandonment of drill should be recommended.

3. That the researcher, before beginning his study, should square himself on basic considerations, such as:

—How useful is this knowledge?

—Should it be taught for drill mastery?

—If it is appropriate only for problem or appreciation teaching, should the drill method of checking be used?

4. Considerations of drill mastery are not the only considerations. There are broad curricular considerations. There are considerations involving right teaching methods. The student who forgets entirely the broader considerations involved in the school program is a clear illustration of the man who redoubles his efforts after forgetting his aims.

Those working in the field of arithmetic have too frequently regarded all phases of arithmetic as drill; they have too frequently regarded drill as the only method of teaching arithmetic. This puts arithmetic at considerable disadvantage. We need in arithmetic, as in other subjects, discriminating teaching. Those who do research in the field should understand this, and should be equally discriminating.

THE MATHEMATICAL VOCABULARY USED IN SECONDARY-SCHOOL TEXTBOOKS OF SCIENCE

FRANCIS D. CURTIS
University of Michigan

Editor's Note: One function of instruction in any area is the establishment of basic concepts. The author indicates some of the more important concepts of secondary school science.

PUBLISHED investigations dealing with the special vocabularies of the branches of mathematics taught in secondary schools have been relatively numerous during the past three decades. Prominent among such studies are those by Pressey,[1] Thorndike[2] Remers and Grant,[3] Georges,[4] and Kunkel.[5] The relationship between the vocabularies of mathematics and science, has, however, received relatively little attention from investigators. Always important, this relationship has recently taken on added significance as a result of the formulation of pre-induction and post-induction courses and materials designed to facilitate the training of men for armed service. More evidence, therefore, with respect to what mathematical terms are commonly found in textbooks of high-school science is needed both as a guide to the teaching of this vocabulary in the courses in mathematics taken prior to or coincident with those in science, and as a means of integrating more closely the work of mathematics and science in junior and senior high schools.

STATEMENT OF THE PROBLEM

The purpose of this investigation[6] was to determine which "difficult" words found in textbooks of high-school science are mathematical terms.

[1] Pressey, Luella C. "The Determination of the Technical Vocabulary of the School Subjects," *School and Society,* XX (July 19, 1924), 91–96.

[2] Thorndike, E. L. *The Psychology of Algebra.* New York: The Macmillan Company, 1923. Pp. 260–273.

[3] Remmers, H. H. and Grant, A. "The Vocabulary Load of Certain Secondary School Mathematics Textbooks." *Journal of Educational Research,* XVIII (October, 1928), 203–210.

[4] Georges, J. S. "The Nature of Difficulties Encountered in Reading Mathematics." *The School Review,* XXXVII (March, 1929) 217–226.

[5] Kunkel, Paul Vance. *The Vocabulary of High School Algebra.* Allerton, Pa.: H. Ray Haas and Company, 1939.

[6] This study by Herbert Hannon is the final one in one group of 34 unpublished master's theses which were completed over a number of years in the author's "Seminar in Secondary-school Curriculum." Hannon's study is a synthesis and further analysis of data from the other thirty-three theses, each of which included a word-by-word analysis of the entire content of one widely-used textbook of high-school science, and

THE PROCEDURES AND FINDINGS

Since no definition of a mathematical term could be found of which practical use could be made in this study, the first task was to prepare a set of criteria to be used in the selection of mathematical terms. The criteria of a scientific term which had been formulated and in successive years had been modified through repeated use by student-researchers in the "Seminar in the Teaching of Science" at the University of Michigan, were revised by the investigator (Hannon) with the help of a university specialist in the teaching of mathematics. These criteria were subsequently modified as their shortcomings became manifest during the progress of the investigation. The final list which follows, though perhaps neither complete nor composed of criteria which are in every respect mutually exclusive, proved satisfactory:

1. A word or a phrase is considered to be a mathematical term if it conveys to an individual trained in mathematics a connotation which it would not convey to a person not so trained. Examples: *adjacent, radical, factor.*

2. A word or a phrase representing a relationship which may be expressed by means of a formula or an equation is considered to be a mathematical term. Examples: *pi, binomial theorem.*

3. A word or an expression designating a magnitude, a quantity, or a number, or one having to do with the properties of form and space (as studied in geometry) is considered to be a mathematical term. Examples: *horsepower, area, million, perpendicular, hexagon, plane.*

4. A prefix having a mathematical connotation is considered to be of itself a mathematical term, even when attached to a root having no mathematical meaning. Examples: *pent* in *pent*oxide, *mono* in *mono*basic or in *mono*rail.

5. The name of a unit of measure is considered to be a mathematical term, but, with the exceptions stated later in this paragraph, the name of

---

all of which together present a study of more than five million running words contained in 9 textbooks of physics, 8 of chemistry, 10 of general science, and 6 of biology. Most of the studies from which Hannon secured his data provided also some of the data which were considered earlier in the report of a synthesis of one hundred investigations of the vocabulary of high-school science.*

* Curits, Francis D. *Investigations of Vocabulary in Textbooks of Science for Secondary Schools.* Boston: Ginn and Company, 1938.

The master's theses which served as sources of data for this study are those by George Bachur, T. F. Bartlett, C. E. Borberg, C. E. Bush, Dean Challis, C. C. Chase, R. S. Collins, W. A. Dexter, Dorothy Dorsey, W. J. Elliott, W. I. Fiedler, C. E. Hanshew, K. R. Houston, W. J. Hyink, H. W. Kollen, G. F. Kyser, G. I. Larkin, Russell McDougal, V. W. Miles, Grayson Newark, J. I. Niergarth, William Ornstein, B. H. Phipps, W. Y. Pohly, M. A. Ruisanchez, F. R. Schauer, William Schultz, R. T. Schroeder, George Spotts, W. B. Toothacker, E. J. Wearne, L. G. Wilkinson, and A. W. Wilson.

the instrument for measuring the unit is not. Thus *volt* and *candle power* are mathematical terms but *voltmeter* and *photometer* are not. The names of Instruments used for measuring units used in a srtictly mathematical sense, however, as *yardstick, meter stick,* and *ruler,* are considered to be mathematical terms.

6. An expression a full comprehension of which can be gained only from considerable experiencing with numbers and quantities, and which cannot be completely understood by the mere learning of a definition, is considered to be a mathematical term. Examples: *ratio, equation.*

As is evident in these criteria, a mathematical term may consist of two or more words each of which alone would have either no mathematical connotation or a connotation different from that of the entire expression. Examples: *inverse ratio, mixed fraction, greatest common divisor.*

These criteria were applied by the investigator to the lists of "difficult" words secured from the thirty-three theses. In these investigations, a "difficult" term was defined as one which does not occur in the first six thousand words of Thorndike's twenty-thousand-word list.[7] This definition was established from the results of 45 investigations of vocabulary difficulty encountered by groups of junior- and senior-high-school pupils when reading textbooks of science.[8]

It will be noted that a limitation was imposed upon the present study by the necessity for deciding which of the words and phrases in the various separate lists were mathematical terms when those words and phrases were out of context. The impossibility of determining with assurance whether any one of a considerable number of these words did or did not possess mathematical connotations in their original contexts is at once obvious. For example, *table* and *series* might or might not be mathematical terms depending on the way in which each had been used. In every such case of doubt the investigator decided arbitrarily to consider that the word was a mathematical term in order to insure that such errors as might thus be made would be on the side of including in the resulting lists all words which could *defensibly* be judged to be mathematical terms rather than of excluding any terms which might logically, but could not certainly, be said to be mathematical.

[7] Thorndike, E. L. A Teacher's Word Book of the 20,000 Words Found Most Frequently and Widely in General Reading for Children and Young People. New York: Bureau of Publications, Teachers College, Columbia University, 1931.

[8] Curtis, *op. cit.*, Chapter III.

Each word or phrase thus judged by the investigator to be actually or possibly a mathematical term was listed by him and the number of times each term had been found in each of the thirty-three textbooks was recorded from the source studies. These terms arranged in four lists respectively representing the textbooks of physics, chemistry, general science, and biology, in which the terms had originally been found, were then submitted to a second person, trained extensively, like the investigator, in the field of mathematics. This person applied the criteria and crossed from each list every word or phrase which, in his opinion, could not logically be considered as likely to be a mathematical term. The resulting four lists which now included only the terms approved for inclusion by both judges, were submitted for further refinement to the university specialist in the teaching of mathematics who had assisted the investigator in refining the criteria of a mathematical term. This third judge in turn applied the criteria and deleted from the remaining terms all those which he thought could not justifiably be considered as mathematical terms.

Table I presents data relative to the prevalence of mathematical terminology in thirty-three textbooks of science written for highschool pupils.

TABLE I

DISTRIBUTIONS OF MATHEMATICAL TERMS IN HIGH-SCHOOL TEXTBOOKS OF PHYSICS, CHEMISTRY, GENERAL SCIENCE, AND BIOLOGY

Textbooks of	Average Number of Different Difficult Words per Book	Average Number of Different Difficult Mathematical Terms per Book	Average Percentage of Different Difficult Words that Are Mathematical Terms
Physics	1,407	159	11.3
Chemistry	1,901	90	4.7
General Science	2,109	103	4.9
Biology	3,019	81	2.6

It will be noted in Table I that much the largest average number of different difficult mathematical terms was found in the textbooks of physics and the smallest in the textbooks of biology. It will be noted further that the average percentage of different difficult words that are mathematical terms in the textbooks of physics was more than four times that in the textbooks of biology and more than twice that in the textbooks of chemistry and general science.

Table II presents two types of evidence indicating the relative importance of these mathematical terms in the study of the various branches of high-school science: the average frequencies with which the terms appear in the textbooks analyzed and the terms which were found in half or more of the textbooks of each branch of science. It should be kept in mind, however, that many more mathematical terms than those listed in this table appeared sporadically in the lists of difficult words taken from the reports of the analyses of the thirty-three textbooks. It should also be remembered that the terms listed in Table II are those that either appear in or above the seventh thousand-word level of the Thorndike twenty-thousand word list or are not sufficiently common in the literature which Thorndike analyzed to earn places in that list; and that therefore the terms in Table II should be supplemented with the mathematical terms found in the first six thousand-word levels of the Thorndike list.

TABLE II

THE DIFFERENT DIFFICULT MATHEMATICAL TERMS WHICH OCCURRED ON AN AVERAGE THREE OR MORE TIMES PER BOOK OR WHICH OCCURRED IN HALF OR MORE OF THE TEXTBOOKS OF PHYSICS, CHEMISTRY, GENERAL SCIENCE, AND BIOLOGY

Mathematical Term	Average Number of Times per Textbook of			
	Physics	Chemistry	General Science	Biology
accumulate	*	3.1*		8.5*
accuracy	4.9*	*	*	*
accurate	19.4*	8.9*	11.2*	8.3*
acute	*			*
adjacent	6.9*			3.7*
algebraic	*			*
alpha	8.5*	7.4*		
ampere	106.4*		9.5*	
analysis	*	*	*	*
analytical		*		
analyze		12.6*	*	3.8*
angular	*			
approximate	16.5*	15.6*	14.1*	8.3*
approximation	*			
arc	27.1*	10.9*	4.5*	*
assumption	*	*		
auxiliary	*			
beta	5.1*	5.1*		
bi-		16.3*	*	
billion	*	*	*	*
billionth		*		
calculate		*		
calculation	30.7*	33.8*	7.1*	4.8*
calibrate	7.2*	4.6*	*	
	4.1*			

† Table II is read thus: The mathematical term **accumulate** is used on the average 3.1 times per book in the 8 textbooks of chemistry analyzed; it appeared 8.5 times per book in the 6 textbooks of biology and it was found fewer than three times per book in the 9 textbooks of physics and the 10 of general science. It appeared in at least half the textbooks of physics, of chemistry, and of biology, but in fewer than half the textbooks of general science.

TABLE II—Continued

Mathematical Term	Average Number of Times per Textbook of			
	Physics	Chemistry	General Science	Biology
[illegible]	43.5*	21.4*	33.4*	15.5*
[illegible]		*		
[illegible]	4.4			
[illegible]		6.0*		
[illegible]				*
[illegible]				
[illegible]	29.6*	3.6*	12.7*	
[illegible]	108.0*	14.0*	5.7*	*
[illegible]	*			
[illegible]	13.4*		*	
[illegible]	33.4*	5.6*		
[illegible]	4.1*			
[illegible]	5.1*			
[illegible]	5.3*	16.8*	7.7*	20.3*
[illegible]	25.9*	*		*
[illegible]	3.3*			
[illegible]	18.3*	*		*
[illegible]	38.6*		4.2*	
[illegible]	*	*		*
[illegible]				3.0
[illegible]				*
[illegible]				
[illegible]	3.9*			*
[illegible]				*
[illegible]				*
[illegible]	15.0*		3.1*	
[illegible]	3.3*			
[illegible]	36.1*	16.8*	7.6*	
[illegible]	3.7*			
[illegible]				*
[illegible]	10.3*			
[illegible]	7.8*			
[illegible]	3.6			3.2*
[illegible]	70.3*	25.0*	12.8*	6.3
[illegible]	4.9	*		
[illegible]	28.5*		*	*
[illegible]	23.4*	4.8*	14.9*	17.2*
[illegible]	10.9*	*	*	*
[illegible]	4.9*	3.0*		7.5*
[illegible]	*	*		
[illegible]	*			*
[illegible]				*
[illegible]			*	8.5*
[illegible]		*		
[illegible]		17.8*		
[illegible]	7.4*			
[illegible]	45.0*	37.3*	47.7*	58.2*
[illegible]	*	*		
[illegible]	59.2*	10.0*	23.1*	11.5*
[illegible]	7.6*			
[illegible]				*
[illegible]			*	
[illegible]	6.0*			
[illegible]	34.0*	3.5*	6.4*	*
[illegible]	14.5*			
[illegible]	3.9*			
[illegible]	7.3			
[illegible]	5.3*	*		9.3*
[illegible]				*
[illegible]				*
[illegible]				
[illegible]	*		*	*
[illegible]	35.0*	208.9*		
[illegible]	3.1			
[illegible]	28.8*	11.9*	*	*
[illegible]	10.8*			
[illegible]	*			
[illegible]	15.4*	5.9*	*	16.7*
[illegible]	19.7*	9.5*	39.0*	26.5*
[illegible]	20.8*		7.6*	*
[illegible]	3.6*			

TABLE II—Continued

Mathematical Term	Average Number of Times per Textbook of			
	Physics	Chemistry	General Science	Biology
focal	17.4*			*
foci	*			
foot-candle	5.1			
formula	23.6*	195.1*	4.3*	3.0*
fractional	*	9.9*		
frequency	119.4*		12.9*	
gamma	5.6*			
generate	13.2*	9.5*	16.6*	*
geometrical	*			
geometry	*			
gram	127.5*	157.0*	16.2*	4.7*
graph	3.4		*	3.0*
graphic	*	*		
graphical	3.1			
halve				5.0*
hept-		*		
hex-		3.4*		
horizontal	40.0*	5.1*	7.1*	5.2*
horsepower	7.8*		4.5*	
hypotenuse	*			
hypothesis		5.3*	5.0*	
identical	*	5.1*		3.8*
identity				*
imaginary	*		*	
improper		*	*	3.5*
incidence	15.8*		*	
incorrect		*	*	
indefinite			*	
indirect	3.3*	*	3.5*	7.2*
indivisible		*		
intercept	*			
intersect	*		*	
intersection	3.0*			
inverse	15.8*	*		
inversion		*		
irregularity			*	
isolate			*	
joule	10.1*			
kilocycle	4.4*		*	
kilogram	22.0*	*		
kilometer	4.9*		3.1	
kilowatt	16.6*		5.8*	
lateral	*		*	
latitude	*		8.5*	8.5*
linear	11.4*			
liter	13.4*	128.6*	*	
literal			3.0*	*
longitude	*		7.5*	
magnitude	15.3*			
mathematical	4.8*	*	*	*
mathematician	*		*	
maximum	23.6*	*	*	5.0
metric	16.1*		*	*
microfarad	*			
mil	8.9*			
mile	3.6			
milligram	*	*		
millimeter	12.9*	7.9*		
millionth	*	*		
minimum	5.4*	*	3.7*	*
minus	4.4*		*	
multiple	9.8*	6.3*	*	
negligible		*		
numerical	8.9*	4.6*		
oblique	5.8*		*	
oblong				*
oct-		*		
ohm	82.3*		4.0	

TABLE II—Continued

Mathematical Term	Average Number of Times per Textbook of			
	Physics	Chemistry	General Science	Biology
orbit		8.6*		
originate				*
oval			*	4.7*
parabola	*			
parabolic	*			
parallelogram	13.0*			
parenthesis		*		
pent-		6.3*		*
percent	4.9*	75.9*	10.4	38.7*
percentage	4.2*	60.4*	11.0*	5.0*
perpendicular	21.3*		4.3*	
plus		3.5*	3.2*	*
poundal	4.9*			
precision	*			
primarily			*	
prism	21.8*	5.4*	6.2*	
prismatic		6.4*		
probability			*	
projection	9.4*		4.8*	10.0*
proportional	60.2*		*	
proportionate				*
pyramid		*	*	*
quadruple				*
quantitatively		*		
quotient		*		
radical		43.5*		
radii	*			
radius	20.3*		*	*
ratio	34.7*	16.0*	3.9*	7.8*
reciprocal	*			
rectangular	5.4*			
relationship	*	3.0*	4.9*	13.7*
resultant	31.0*			
rhombic		5.4*		
rotate	52.5*	6.0*		*
rotation	16.8*		5.2*	3.5*
secondary	44.7*	*		3.5*
segment			*	*
semi-	3.1	*		
semicircular	*			
significant		*		
simultaneous	4.8*	*		
similarity	*	*	*	7.0*
simplify		3.0*		
sine	4.1*			
spherical	22.9*		*	3.8*
spiral	4.6*	*	3.7*	5.3*
statistics				*
substitution		5.0*		
superimpose	*			
supplement			*	*
surplus			*	
symmetrical	*			
tangent	3.8*			
tetra-		15.0*		
thousandth	4.2*			
tri-		32.0*		
triad	*			
triangular			*	*
triple	3.2*			
variable	12.4*	5.3*	4.4*	*
variation	16.8*		13.0*	14.5*
verify	3.4*	*		*
volt	103.9*	*	9.7*	
uni-		5.0*		
watt	55.3*		8.7*	
yardstick			*	

FACTORS WHICH AFFECT THE SALARY OF HIGH SCHOOL TEACHERS OF COMMERCIAL SUBJECTS IN MISSOURI

HARRY PELLE HARTKEMEIER
University of Missouri

Editor's note: Most professional workers have at least some general idea about the factors contributing to differences in salaries. The author makes this information more precise for the salaries of commercial teachers in the state of Missouri by the statistical analysis provided in the paper to follow.

SEX, length of employment, and size of the high school are the three factors which are investigated in this study of the variation in the salaries paid high school teachers of commercial subjects in Missouri. It is not possible to study the salaries of all teachers of commercial subjects, for the Missouri School Directory, 1939–40, does not contain information about all schools. The return of the questionnaire sent out by the superintendent of schools for the state of Missouri is not mandatory and some questionnaires were not returned at all. Furthermore the information concerning salaries was omitted more frequently than any other type of information on the questionnaires that were returned. In view of this fact, the study must be considered an analysis of a sample of the data for the whole state. While it is true that the data available do not constitute a strictly random sample from the entire state, the data are quite as satisfactory as most data obtained by the questionnaire method.

Not all of the available data were used, for the statistical technique employed (the analysis of variance with three criteria of classification) requires that the frequencies in the various classes be proportional. After a study of all of the data available the frequencies shown in Table I were decided upon. It was impossible to obtain the desired number of observations when the study was restricted to high school teachers who taught only commercial subjects, so it was necessary to include high school teachers whose major work is the teaching of commercial subjects. Commercial subjects include primarily bookkeeping, typing, and shorthand, and, less frequently, commercial arithmetic, commercial law, and business English.

This study was restricted to first class high schools, but this is not a serious limitation due to the fact that second and third class high schools usually do not offer commercial work. The classification labelled "large"

schools does not include schools which had thirty or more high school teachers, for all of these schools omitted salaries from the answers to the questionnaire.

The frequencies shown in Table I come as near to being proportional to the populations from which they were drawn as can be expected for the size of the sample selected. Actually there are slightly more than twice as many women commercial teachers as men. The average "tenure" of high school teachers is between two and three years. While the frequency of observations (number of teachers) for small schools may seem too small, it should be remembered that many teachers in small high schools are part-time teachers. This study is concerned only with full-time teachers.

TABLE I

NUMBER OF OBSERVATIONS SELECTED

Years in System	Sex					
	Male			Female		
	Size of School			Size of School		
	Small	Average	Large	Small	Average	Large
1	4	8	2	8	16	4
2	4	8	2	8	16	4
3	4	8	2	8	16	4
4	2	4	1	4	8	2
5	2	4	1	4	8	2
6	2	4	1	4	8	2
Totals	18	36	9	36	72	18
	63			126		
	189					

The effect of the length of employment in the present position after six years is not included in this study because most regular programs of systematic raises in salary or salary schedules end after the fifth or sixth year. Very few teachers of commercial subjects have remained in the same position more than six years.

The analysis of variance with three criteria of classification and proportional sub-class frequencies was applied to the data to obtain the results shown in Table II.

The variation within classes is the best measure of random variation because it is not affected by the three principal components being studied. Sex, size of school, and length of tenure are all held constant in the measurement of the variation within classes. However, when the variation associated with each one of these three factors is compared with the variation within classes, it can easily be seen that the effect of each one of the factors is highly significant or reliable. The first three observed F ratios shown in Table II are all much larger than the ones which would occur 1 percent of the time due to chance alone.

TABLE II

ANALYSIS OF VARIANCE OF SALARIES OF MISSOURI HIGH SCHOOL TEACHERS OF COMMERCIAL SUBJECTS

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Ratios		
				Observed	5%	1%
Sex	1	239, 763	239, 763	24. 4	3. 91	6. 81
Size of School	2	423, 056	211, 528	21. 6	3. 06	4. 75
Years in Present Position	5	564, 689	112, 938	11. 5	2. 27	3. 14
Two-Way Interactions						
Years and Size	10	240, 115	24, 012	2. 45	1. 89	2. 44
Sex and Years	5	85, 901	17, 180	1. 75	2. 27	----
Sex and Size	2	18, 459	9, 230	----	----	----
Three-Way Interaction						
Sex, Size and Years	10	151, 394	15, 139	1. 54	1. 89	----
Within Classes	153	1, 501, 642	9, 815	----	----	----
Total	188	3, 225, 019				

If the reader is not familiar with the use of the F ratio to test the significance or reliability of the effect of a factor on a variable, the following brief illustration may help.[1] The mean square 239,763 is a measure of the amount of variation in salaries caused by (or associated with) the sex of the teacher. In order to discover whether this is a reliable value or merely the effect of random fluctuations a statistician usually calculates the F ratio. This is the ratio between the measure of the effect of sex, 239,763, and a measure of the effect of purely random factors on the salaries of high school teachers of commercial subjects. The best available measure of the effect of

[1] For a more detailed explanation the reader should consult Snedecor, George W., *Statistical Methods, Collegiate Press,* 1938, pages 179–188, or Hartkemeier, Harry Pelle, *An Introduction to Managerial Business Statistics,* Thomas Y. Crowell, 1943, pages 28–30, 52–53.

random fluctuations is provided by the amount of variation within classes, 9,815. This value is a measure of the amount of variation in salaries paid a group of teachers of the same sex, same size of school, and the same number of years in the present position. This procedure is adopted by the statistician in order to obtain a measure of variation free from the effect of the three principal factors he is studying. The statistician then divides the measure of the effect of sex by the measure of the effect of random events. In this case we divide 239,763 by 9,815 to obtain 24.4. If this ratio had been 1.0 statisticians would not have been able to conclude that sex has an effect upon salaries—that is, an attempt to compare the salaries of teachers of different sexes would yield nothing more than a comparison of the salaries of teachers of the same sex. However, we did not obtain a ratio of 1.0; we did obtain a ratio of 24.4. This shows that the variation in salaries caused by sex is 24.4 times as large as the variation in salaries caused by random events.

It is extremely improbable that such a large ratio as 24.4 could occur due to chance alone—that is, due to an unusual sample from a group of teachers of different sexes who receive salaries that are independent of sex. The probability of encountering an F ratio of 24.4 due to chance alone is much less than 1 percent. This is shown in the fact that an F ratio of 6.81 would not occur oftener than 1 percent of the time due to chance alone. This 1 percent value may be found in prepared tables of the F ratio.[2]

When any F ratio in a sample (see the observed F ratios in Table II) is larger than the F ratio that would occur less than 5 percent of the time due to chance alone, statisticians accept the observed F ratio as reliable proof that the factor under consideration caused (or is associated with) a real difference in the variable being studied (salary, in this case). The F ratio of 24.4 is accepted as reliable proof that sex affects teachers' salaries. In this case the F ratio is larger than the 1 percent value, so statisticians are very sure (still more confident) that sex affects teachers' salaries. If the F ratio had been 6.81 statisticians would be right 99 times out of 100 if they were to say that sex affects teachers' salaries. The observed F ratio is much larger than 6.81, so statisticians will be right much oftener than 99 times out of 100 if they conclude on the basis of the sample analyzed that sex does affect teachers' salaries.

[2] Op. cit.

TABLE III

NUMBER OF OBSERVATIONS AND AVERAGE SALARY

Sex, Size of School

Size of School	Sex		Total
	Men	Women	
Small	18 $1019	36 $917	54 $951
Average	36 $1072	72 $1014	108 $1034
Large	9 $1158	18 $1062	27 $1094
Total	63 $1069	126 $994	189 $1019

TABLE IV

NUMBER OF OBSERVATIONS AND AVERAGE SALARY

Sex, Years in Present Position

Years in Present Position	Sex		Total
	Men	Women	
1	14 $1019	28 $931	42 $960
2	14 $1042	28 $981	42 $1002
3	14 $1084	28 $1004	42 $1031
4	7 $1028	14 $983	21 $998
5	7 $1030	14 $1031	21 $1031
6	7 $1275	14 $1095	21 $1155
Total	63 $1069	126 $994	189 $1019

A study of the averages shown in Tables III, IV, and V reveals that on the average men are paid more than women and larger salaries are paid by larger schools. It is not true, however, that teachers who have held the same position for a longer period of time receive the higher salaries. Up to and including three years, the length of time in the same position helps

increase the salary; but for both sexes in all three sizes of school the teachers who held their positions four years receive less than the teachers who held their positions three years. This may be due to the fact that teachers who can move on to better positions after two or three years do so, leaving the lower salaries to those who cannot move. Possibly the money is used to retain the good teachers for two or three years, and funds are insufficient to retain them longer. In any event, the variation in salaries associated with years in present position is significant or reliable.

TABLE V

NUMBER OF OBSERVATIONS AND AVERAGE SALARY

Size of School, Years in Present Position

Years in Present Position	Size of School			Total
	Small	Average	Large	
1	12 $904	24 $979	6 $995	42 $960
2	12 $908	24 $1042	6 $1025	42 $1002
3	12 $949	24 $1033	6 $1183	42 $1031
4	6 $934	12 $1000	3 $1119	21 $998
5	6 $947	12 $1080	3 $1005	21 $1031
6	6 $1158	12 $1113	3 $1317	21 $1155
Total	54 $951	108 $1034	27 $1094	189 $1019

The reader may say that he would have guessed that the three principal factors were important without this proof. It is comforting to know that your beliefs can be proved by quantitative analysis, but the three two-way interaction mean squares yield information that would be very difficult to obtain without the analysis of variance. It answers such questions as these:

1. Does the difference in salary due to sex remain constant or gradually vanish as the teacher gains experience?
2. Do large and small schools discriminate against one sex to the same extent?
3. Does the effect of the number of years in the present position change from one size of school to another?

The last question is answered by the significant interaction[a] between the size of school and the number of years in present position. Values as large as or larger than this one would not have arisen oftener than one time in a hundred due to chance alone if there were no change in the effect on salary of years in present position as we compare different size schools. It appears that larger schools are more apt to have (and use) schedules of systematic raises in salary.

The small and non-significant interaction between sex and the size of school indicates that there is no significant change in the sex discrimination in the setting of salaries as we compare different size schools. The interaction between sex and years is not large enough to be significant. It provides some evidence of a change in the sex discrimination when the teachers have held the same position for a different number of years, but the observed interaction could have arisen due to chance alone oftener than five times in a hundred and we would have to have better evidence before we could conclude that time would erase the sex discrimination in salaries.

[a] The statistical term *interaction* does not necessarily mean that a change in the size of either factor has a causal effect upon the result of a change in the size of the other factor so far as the variable (salary) is concerned. The author thinks, in this case, there is a causal relationship.

SOME MAJOR PROBLEMS IN EDUCATIONAL MEASUREMENT

BENJAMIN S. BLOOM
University of Chicago

Editor's Note: After the splurge of interest of the teens and early twenties in the measurement of educational outcomes many persons thinking that all had been learned turned their attention to new and greener pastures. The author suggests some new problems in this area.

IN A RECENT article, A. E. Traxler* suggested certain problems arising out of attempts to provide more adequate measures of some of the less easily measured aspects of education that seemed to him of major significance. The problems suggested by Mr. Traxler seem to the present writer to give primary emphasis to the mechanical aspects of testing. The major problems he mentioned may be summarized as:

1. Securing cooperation of test users in test making because (a) they must be "convinced that the test is sound" and (b) "the experience will help them to appreciate the high degree of intelligence and work which goes into the making of a good item."
2. "Obtaining comparability among tests in different fields and different tests in the same field, and . . . getting test users to understand the basis of comparability."
3. Identifying the aspects of each testing area which are not highly correlated, constructing long enough tests of these aspects to secure acceptable reliability, and persuading test users to allow enough time for testing to provide a dependable basis for diagnosis and guidance.
4. Determining "the relationship of factor scores to future educational and vocational success."
5. Determining "the adjustment between test booklets and answer sheets."

The selection of these problems as the primary concern of test makers assumes that adequate testing procedures have already been worked out, hence all that remains is further to perfect these instruments and to make certain that consumers of tests make adequate use of them. In this case test makers need make only minor revisions in their tests and can devote most of their time to statistical studies of test results. They could strive to realize the dream of "setting up some common national basis of reference such as

* Traxler, Arthur E., "Problems Arising Out of the Attempt to Apply Improved Measurement Techniques to Education and Guidance. *Journal of Educational Research,* XXXVII, (September, 1943), pp. 14–18.

'the standard million' and of the day when the results of all tests might be expressed in terms of norms derived from a stable group of that kind."

But, test making has not reached this millenium. We are still confronted with many problems which need to be solved—problems which undoubtedly have a higher priority rating than those listed above.

One major problem, the problem which is most general in scope, is the incorporation of testing as an integral and intimate part of the entire learning process. Teachers too frequently regard tests as instruments to be used, results noted, and then filed permanently. Standardized tests are too often regarded as devices to be used because they are educationally fashionable. Such devices are to be made by experts who utilize "high-powered" statistics and who have other mysterious and devious methods which are beyond the comprehension of the average teacher.

If progress is to be made in the field of education, teachers must see their job clearly. They must know their own purposes and goals, they must know when they achieve these goals, and they must know where and why they fail. To do this, they must set up their objectives, measure initial status of students with respect to these objectives, teach, measure final status of students, analyze results to determine success and failure; then they must repeat the cycle over and over again, revising objectives, revising instructional procedures, revising measurement devices—providing a continuous program of educational improvement.

To strive to achieve certain objectives, the teacher himself must cherish these goals—they must be his in every sense of the word—and they must be stated, not to convince others, but to furnish working guides for the total learning process. To understand an objective clearly, the teacher must be able to express it in operational terms, in terms of student behavior and in terms of test items. To have confidence in the measurement devices, enough confidence to make changes on the basis of measurement results, the teacher must have a thorough understanding of the tests.

Here then is the major problem for the test expert, to help teachers measure each of the objectives they are striving to attain, whether tangible or intangible.

Closely bound up with the above problem, in reality a subdivision of it, is the more specific problem of identifying the major objectives of instruction. This is the first step in achievement test construction. The purposes of

educational institutions and of the courses which make up the educational program have, for the most part, been stated in terms so general that they are all things to all men. Each person can and does interpret them in any manner he pleases. As such they offer little guidance to the individual instructor or test maker. The test maker can aid teachers in clarifying their objectives, in stating them in terms of student behavior, in stating them in terms of the tasks the students should be able to perform if the objective is to be fully attained. Through several years of persistent attack on this problem, the test experts and teachers should be successful in mapping out the range of objectives in each course of instruction now offered in our educational institutions and in stating them in such terms as to offer guidance for all who state or select them.

As objectives of instruction are identified, test makers must aid teachers in devising techniques for measuring their achievement or lack of achievement. This is a tremendous problem and the test experts will have to invent measuring techniques which extend beyond the paper-and-pencil procedures which form such a large part of our present stock in trade. Some especially difficult aspects of this problem are those arising from objectives involving personal and social values.

As instruments are made to measure each of the important objectives of instruction, the test expert and the teacher face the problem of simplifying the reporting of scores and of securing the maximum use of tests and test results. Most test makers and users can now point to shelves and files containing many thousands of tests which have been administered and neatly filed. These tests represent many hours of construction and scoring time, and many thousand hours of student time. Each hour of such time should serve a distinct purpose and should be put to maximum use in educational and individual diagnosis. Tests should be used to motivate students and instructors and to guide them in various ways. We cannot justify the administration of a test and the use of valuable student time merely to satisfy the idle whim of an individual in an educational administrative position.

Another problem which test experts must face are the many research questions which arise from solutions to the above problems. Educational research needs to be designed with the care which now characterizes research in the biological and physical sciences. From such careful research designs, the test maker should determine the effects of instruction on the relationship

between measures of various objectives. He must determine the extent to which each of the objectives of instruction are retained, once students have achieved them. He should assist teachers in finding those instructional techniques which facilitate or retard the student's progress toward the achievement of various objectives. And, he should conduct research which will determine the various ways in which people achieve various educational goals—through maturation, through informal educational experiences, and through experiences which can be described as educational only by the broadest definition of this term.

These are the types of problems that test makers should put first on their priority lists if educational measurement is to develop as a primary element of effective education.

DIRECTIONAL TESTS FOR EDUCATIONAL GUIDANCE

FORREST H. KIRKPATRICK
Bethany College, Bethany, W. Va.

Editor's Note: The war is giving new emphasis to guidance activities. The author offers certain definitions and a plea for aptitude measures that measure latest learning abilities.

THERE is considerable confusion as to the use of aptitude testing in terms of different types of work or study. Part of this confusion arises because educational and vocational problems can hardly be differentiated, and part because of an overlapping of objective measures of *achievement* with those of *aptitude*. Thus, the college dean or guidance officer, influenced by the "round peg in square hole bogie" of vocational guidance, and lacking a clear conception of "educational aptitude", may prematurely emphasize the choice of a career at a time when emphasis should rather be placed upon curricular planning appropriate to the student's particular kind of readiness to learn.

As an example, we may compare "vocational aptitude" for accounting with "educational aptitude" for mathematical studies. Accountancy is an admirable profession, offering excellent opportunities for those possessing the attributes essential to success therein. Among these is at least one phase of mathematical ability. The same basic learning power, or educational aptitude in respect to dealing readily with figures and quantitative concepts, which is demanded for qualification for the C.P.A. degree, might find equal scope in other fields. Astronomy, actuarial work, statistical method, and certain operations in physical science readily come to mind. The important point is first to identify and cultivate mathematical aptitude where it exists.

The term aptitude itself is often so loosely employed that it must be here defined. The accepted use of this word among students of mental measurements is substantially that given in Warren's *Dictionary of Psychology*.[1] An aptitude is there defined as a

"condition or set of characteristics regarded as symptomatic of an individual's ability to acquire training in some (usually specified) knowledge, skill, o set of responses such as the ability to speak a language, to produce music, etc.'

[1] Howard C. Warren, ed., *Dictionary of Psychology*. New York: Houghton–Mifflin Company, 1934.

Two points especially are to be noted in this definition. In the first place, aptitude is differentiated from skill. Skill is the ability to perform some given responses at a given time; aptitude is the ability to acquire skill under appropriate conditions, regardless of whether those conditions have arisen or not. In the second place, the definition does not involve any assumption as to whether aptitudes are acquired or innate even though it is important theoretically. This usage is essentially the same as that outlined by W. V. Bingham in *Aptitudes and Aptitude Testing*.[2]

Within the field embraced by this definition, colleges and universities are more concerned as to a still more limited area, which, for convenience, may be called the area of *educational aptitudes*. Their concern is in terms of the students' abilities to acquire knowledge and skills demanded for specific curricula. An interesting parallel exists in methods of approach to the testing technique between primarily vocational and the needed educational instruments. For instance, finger dexterity tests to measure "teachability" or apprenticeship-promise are administered before assigning girls to vestibule school training in radio tube mounting operations. It is clear that aptitude tests, as now understood, have been utilized more extensively by industry at the *skill level* than by colleges at the higher levels.

As far as colleges are concerned, the ultimate utilization that is made of students' aptitudes, after further training in a particular field, is surely less important than is prompt recognition and development of this learning capacity *per se*. It is likely that vocational guidance will take care of itself in time, if intelligent planning has laid the proper foundation for personal growth and intellectual development. Narrow specialization based on specific vocational guidance is not sound. The choice of any particular occupation should not be prematurely or narrowly determined while the educative process is still in flux.

From this point of view, the present state of aptitude testing in colleges is far from adequate or helpful. Well validated achievement tests or other measures of learned materials are not of any great use as aptitude identifiers, and they often prove misleading. Resulting data fail to suggest any promise for fields of study to which a student has not been exposed. Hence, neither the student nor the guidance officer can have a suggestion as to other fields

[2] Bingham, Walter V., *Aptitudes and Aptitude Testing*. New York: Harper and Brothers, 1937. p. 17.

which offer scope to intellectual powers of a different type. Achievement tests throw some light on what *he might* have acquired under other educational circumstances, or from other courses than those already experienced, but they are neither prognostic nor analytical.

A similar criticism might be applied to scholastic ability tests as measures of capacity to learn. It is assumed that scores on scholastic ability—or intelligence—tests indicate the extent of one's ability to achieve in most abstract fields. Such tests are supposed to measure abstract intelligence, but they are usually validated in terms of the language-academic curriculum with all the limitations of usefulness that such validation implies.

The difficulty in discovering specific and valid educational aptitudes is further aggravated by factors at work in the extra-college environment. There is always the danger that some combination of primarily subjective circumstances and opinions may press an individual towards some course of study for which his mental make-up is not well suited. Often a student thus misdirected may not realize the presence of relative talents and limitations for a particular educational area until rather late for optimum development. In the extreme instances—which are not few—such potentialities may never be explored, with some frustration rather than fruition as the result. *There is a clear need for some kind of aptitude measurements which will discover latent learning abilities of a specialized sort*—and such abilities that might not otherwise be discovered. We need aptitude tests that have directional significance.

There is no intention here to minimize the difficulty of fulfilling this need. There is very little with which to start. Yet, research-minded educational psychologists who choose to meet the challenge will some day develop educational aptitude tests to the degree that they have intelligence and achievement tests. It is likely that they will find for them equal, if not greater, acclaim in the field of higher education and guidance.

EDITORIALS

THE SYNTHESIS OF SCIENTIFIC STUDIES

IN THIS issue of the *Journal* there are included several articles on problem solving in arithmetic. They illustrate a helpful approach to the integration of research dealing with a central problem. In two of the articles the procedure involves the comparison of results of tests administered to good and poor achievers in order to determine characteristic differences between the two groups of pupils. The traits studied in one investigation were largely related to the computational aspects of problem solving, while in the other skills related to reading were analyzed and compared. It is significant that in both studies traits were found in which there was not a significant difference between good and poor achievers while there were other elements in which reliable differences were found. A synthesis of the two studies indicates that there was a reliable difference between the two groups in general reading ability but that such differences were not found when a comparison of ability was made in specific reading skills such as those measured by Gates Silent Reading Tests, all of which deal with the reading of literary material rather than content that is essentially arithmetical and quantitative in content. On tests of reading skills such as are involved in the solution of problems, for example, finding the key question to a problem, or answering questions based on the steps in solving a problem, reliable differences between the two groups were found. Similarly there was not a reliable difference between the two groups in general information, but there was a reliable difference between them in their knowledge of arithmetic information such as understanding of social uses of number, and an appreciation of the general application of quantitative procedures. In general there was a reliable difference between the two groups in various computational skills, such as estimating answers to examples and problems, manipulation of number operations, and seeing relations in number series. There were also reliable differences between the two groups in a variety of psychological factors, such as general reasoning ability, and general language ability.

The general conclusion to be derived from these studies is similar to that which has been reported by a number of other investigators of reading in content fields. It seems evident that the reading skills required in various subject-matter fields vary from one area to another. It is clear that the teacher of arithmetic who wishes to develop ability of her pupils to solve problems must see to it that the essential reading skills required are developed through well planned guidance. It may be assumed that if pupils who evidence weakness in problem solving are given practice in applying the skills in which superior pupils excel there is likely to be a considerable improvement in their ability to solve problems. This assumption can be established by the kind of experimental work reported by Johnson in one of the articles in this number, dealing with the improvement of arithmetic vocabulary. It should be pointed out that in the first two studies commented on above there were differences in the knowledge of general and arithmetic vocabulary between good and poor achievers, but that these were significant only at the 5 per cent level. Other assumptions derived from preliminary studies can be checked in the same way.

Another point to which these studies make a valuable contribution is the determination of the areas of ability in which the use of these procedures as a basis of instruction will likely be most useful. It is obvious that a given technique may not be equally valuable for all pupils, although it may be of marked assistance to pupils of certain limited mental levels or chronological ages. The statistical techniques used in these three studies make it possible to determine the regions of significance in which there are reliable differences between good and poor achievers. The application of these technical procedures to the analysis of the data made it possible to extract from the information a number of important conclusions that it would not have been possible to draw if comparisons had been made between measures of central tendency. The method of analysis also enabled the investigators to control mental and chronological ages in comparing the two groups, which made it possible to establish regions of significance. Information of this kind is of great value in assisting teachers to select the kinds of instructional procedures most likely to be of value with pupils having certain definite characteristics that are known. If it can be demonstrated by experiment that the

development of various skills in which good achievers excel poor achievers in problem solving results in improvement of that general ability, then we can assume that the wider use of teaching procedures based on these techniques will lead to a general growth of pupils to solve problems. Basic assumptions can thus be established as described above by preliminary studies, and then they can be checked by subsequent experimental investigation.

LEO J. BRUECKNER,
University of Minnesota.

THE INCREASE IN PRECISION IN EDUCATIONAL AND PSYCHOLOGICAL EXPERIMENTATION THROUGH STATISTICAL CONTROLS

Two major statistical problems confront the research worker after he has designed and carried out his investigation to obtain observational data best suited to answer the question which has prompted the research. The one is the problem of estimation which consists in applying the method of reduction which will give him the best unbiased estimate of the effect which he presumes to exist. The second problem is to determine whether the effect is a real one or a spurious one that may be attributable to random factors. For this determination, application should be made of the specific test of significance appropriate to the situation. A problem associated with tests of significance is the specification of the population for which a generalization is permissible. These problems are subjects of this discussion. In the short space available for their treatment, it is not possible to develop the mathematical foundation upon which the statistical methods are based. The technical account has been given elsewhere.[1] This presentation will attempt to answer the inquiries: (1) Why is a technique to increase precision necessary? (2) What are the essential aspects of the technique? and (3) What assistance does the technique give?

It is possible, of course, and not of infrequent occurrence to carry on an investigation in which no treatment that may be accorded the data will lead to conclusive results. Such is often the case when statistical treatment and the planning of the experiment have not been considered as two aspects of a single problem, viz., the problem of organizing or planning an investigation so as to give observations that will provide a sound basis for drawing inferences with reference to the particular hypothesis under consideration. It is assumed, therefore, in this discussion that, the principles, as they may now be known, of sound investigation have been in operation whether the investigation be based upon the arrangement of an experiment to obtain new observational data or be based upon some other systematic observational program. It is also assumed that the research worker is not guided by the belief that it is good enough if he has reason to believe that

[1] Palmer O. Johnson and J. Neyman. "Tests of Certain Linear Hypotheses and Their Application to Some Educational Problems." *Statistical Research Memoirs*. Vol. I, pp. 57–94. Department of Statistics, University of London, University College. June, 1936.

his estimate is at any rate not an underestimate or that the conventional tools are sufficient because his data do not merit more refined treatment. The implication seems to be that bad data call for bad statistics.

The answer to the question of why a technique for increased precision is required may be approached by considering a common practice in educational or psychological experimentation. In order to secure uniformity in the groups of individuals making up the subjects of an experiment, it is customary to couple individuals on one or more basic characters that presumably bring about homogeneity with respect to learning abilities. While this process serves to increase the sensitiveness of the experimental observations, a considerable practical difficulty is met with in securing representation throughout the range of variation and in discovering basic characters that bear a sufficiently high relation to the criterion to guarantee closely homogeneous groups. Furthermore, there are usually a number of administrative gyves that hinder or prevent freedom of selecting and placing subjects. Pairing, of course, may be carried out subsequent to experimentation which may overcome administrative restrictions but not other limitations. Another difficulty is due to the fact that the difference in the mean effect of the contrasted educational factors—frequently constituting the experimental factor under study—likely depends upon the character or characters used in the pairing process.

Difficulties—not limited to pairing—arise also in the measurement of the criterion itself. A frequent procedure is to estimate growth in learning as (1) the absolute difference between initial and final scores on the tests or measures employed, or (2) as the percentage of gain in terms of initial status or of a theoretical psychological limit, i.e., the total attainable score. In using these methods the initial and final measures are treated as if they were independent and therefore the relationship usually existing is neglected.

The difficulties or limitations enumerated may be overcome or substantially reduced if we seek the answers to these questions not through *a priori* considerations, but in the internal evidence provided by the data themselves. In this sense the experiment may be said to be self-contained. The increase in precision of estimation of the presumed effect is, therefore, obtained through statistical controls and is largely a regression problem. It will be noted that if this procedure is permissible the necessity of matching individuals disappears and also that the estimate of the difference in effect of conditions that constitute the experimental factors is no longer a rigid one

and may be referred separately to different classes of individuals. Furthermore, if the matching of individuals is unnecessary, all available individuals can be used.[2] Every statistician knows how data may be falsified when an arbitrary selection is made among the observational material available.

The series of investigations in the teaching of arithmetic reported in this issue of the Journal illustrates the essential features of the statistical method devised for solving the problems enumerated above.

The method makes it possible to test the significance of the difference between the means of groups under comparison adjusted for the inequalities of the basic characters of matching. Thus, individual pairing is unnecessary and all cases can be used.

When the null hypothesis is rejected, the method makes possible the construction of linear forms in the observational data which provide estimates of the population parameters without bias and with minimum variances.

The method also introduces a new concept in statistical analysis, viz., the region of significance which makes it possible to answer the question: For what systems of values of the basic characters of matching would the difference between the adjusted means of groups of similar size be judged significant at a specified level of probability if judgment is based on the results of the test derived by applying the likelihood criterion? In other words, by this method the experimenter can find out (as illustrated in the researches reported here) for what kind of students as described by the matching variables a difference in the criterion exists. If the region of significance proves to be an hyperbola, for instance, then the region would consist of two parts determining values of the basic characters for which alternatively the one or the other of the two groups under comparison may be said to be superior with respect to the criterion of comparison used.

In conclusion, the technique enables the determination of the region of significance, if it exists, which utilizes the measurable amount of information provided by a body of observational experiences concerning the population for which a generalization is permissible.

PALMER O. JOHNSON,
University of Minnesota.

[2] Even if matching is resorted to, increased precision can be gained by making allowance for inequality of groups by means of the statistical controls recommended

Research Abstracts and Bibliographies

Address research bulletins and other communications to A. S. Barr, Department of Education, University of Wisconsin, Madison 6, Wisconsin.

ABSTRACTS

Accrediting Military Experience. Washington, D. C.: U. S. Civil Service Commission, 1944. 10 pp.

A selected list of references.

A Study of Methods of Changing Food Habits of Rural Children in Dakota County, Minnesota. Nutrition Education Series, Pamphlet No. 5. Washington, D. C.: Federal Security Agency, U. S. Office of Education, 1944. 18 pp.

Discusses importance of studying food habits; purpose of the Dakota County study; plan of the project; general conclusions; activities of the schools and effects of food study and school lunches; and recommendations.

BRANDON, ARTHUR L. *Postwar Education in American Colleges and Universities.* Bloomington, Indiana: American College Publicity Association, Research Division, 1944. 76 pp.

Discusses acceleration and concentration; compulsory military training in the colleges; veterans' education; the postwar college and university; junior colleges; two-year bachelor's degree; adult or continued education; college finances; scholarships and loans; teachers' salaries; postwar salaries; postwar enrollment; postwar building by the colleges; cooperation with business and industry; and intercollegiate athletics.

BROWN, EDWIN J. and O'REILLY, VIRGINIA M. *A Self-Rating Scale for the Elementary Teacher.* Bulletin of Information, Volume XXIV, No. 6. Emporia, Kansas: Kansas State Teachers College of Emporia, 1944. 36 pp.

The author presents a self-rating scale for elementary school teachers with data on its validity and reliability.

Bulletins of the California Test Bureau. Los Angeles, California: California Test Bureau, 1944.

These publications discuss: conducting high school guidance programs; planning the elementary school testing program; identifying the difficulties in learning arithmetic; diagnosis in the reading program; appraising personality and social adjustment; and use of tests and inventories in vocational guidance and rehabilitation.

CARTER, HAROLD D. *Vocational Interests and Job Orientation.* Applied Psychology Monographs, No. 2, of the American Association for Applied Psychology. Stanford University, California: Stanford University Press, 1944. Pp. 85.

Discusses developmental nature of vocational interests; the nature of the concepts involved; the varieties of measuring instruments; the reliability and constancy of interests; the validity and general usefulness of available measures of vocational interests; special problems of

techniques; individual differences in interests; the problem of vocational choice; summary and references.

CHAMPLIN, MILDRED WILDER. *Practicum in Training for Adolescent Leadership.* Hartwick College Studies. Blair, Nebraska: Lutheran Publishing House, 1942. Pp. 24.

The author presents a plan for encouraging student participation in the discussion of social problems and issues.

CONLEY, WILLIAM H., Chairman, Evaluation Committee. *The Wright Junior College Study of Terminal General Education.* Chicago, Illinois: Wright Junior College, 1944. 149 pp.

Presents data on terminal education: the Wright Junior College student; evaluations of growth in knowledge measured by standardized tests; evaluations of growth in the ability to think as measured by two tests of the Progressive Education Association Eight Year Study; evaluations of achievement made by means of Chicago Junior College tests and examinations; evaluations of attitudes, interests, and activities; and summary, conclusions, recommendations.

CRAIG, GERALD S. *Science in Childhood Education.* Practical Suggestions for Teaching, No. 8. New York: Bureau of Publications, Teachers College, Columbia University, 1944. 86 pp.

Presents a discussion on science in the lives of children; attitudes and methods of work to which science teaching should contribute; science in the elementary school curriculum; and knowing and using community resources.

CURETON, THOMAS KIRK. *Guide for Tracing Research in the Health, Physical Education, and Recreation Field.* Reprinted from the *Research Quarterly*, May, 1944.

Contains detailed direction on the sources of material on health, physical education and recreation.

EATON, MERRILL T. *A Survey of the Achievement in Social Studies of 10,220 Sixth Grade Pupils in 464 Schools in Indiana.* Bulletin of School of Education. Indianapolis, Indiana: Bureau of Cooperative Research and Field Service, Indiana University, 1944. Pp. 68.

Presents a comparison of mean scores of various groups on age, length of school year, number of classrooms in building, size of class, number of grades in classroom, size of school, time given to study of social studies, departmentalization of social studies, teaching of history and geography as unified or separate courses, and congressional districts.

Educational Bulletins of the California Test Bureau, 5916 Hollywood Boulevard, Los Angeles 28, California.

No. 1. How Tests Can Improve Your Schools; no. 2. How To Select Tests; no. 3. How To Conduct a Survey; no. 4. Administrative Use of Survey Results; no. 5. Teacher Use of Test Results; no. 6. Basic Testing Program; no. 7. Conducting High School Guidance Programs; no. 8. Planning the Elementary School Testing Program; no. 9. Identifying the Difficulties in Learning Arithmetic; no. 10. Diagnosis in the Reading Program; no. 11. Appraising Personality and Social Adjustment; no. 12. Use of Tests and Inventories in Vocational Guidance and Rehabilitation; no. 13. Suggestions for the

Use of Standardized Tests for Educational Planning, Vocational Guidance, and Placement of Persons Committed to Correctional Institutions; no. 14. The Proper Use of Intelligence Tests; and no. 15. Vocational Guidance for Junior and Senior High School Pupils.

Free copies of these bulletins may be obtained by writing to the above address.

First Annual Report of Governor's Commission on Metropolitan Transportation. Providence, Rhode Island: 14 College Street, 1944.

The Governor's Commission on Metropolitan Transportation was formed a year ago to obtain the maximum of transportation conservation and to maintain the most efficient transportation services for essential workers as was possible in wartime.

GREENLEAF, WALTER J. *Teachers are Needed.* Vocational Division Leaflet, No. 14. Washington, D. C.: Federal Security Agency, U. S. Office of Education, 1944. Pp. 25.

Discusses the need for teachers; teaching as a wartime service; career information for the counselor to relay to students; characteristics of successful teachers; personality record for prospective teachers; identifying characteristics suitable for prospective teachers; interviewing teacher candidates; and suggestions for the counselor.

HECK, ARCH O. and others. *Education of Exceptional Children and Minority Groups.* Review of Educational Research, XIV, No. 3. Washington, D. C.: National Education Association of the United States, 1944. Pp. 86.

Discusses general problems and philosophy and administration in the education of exceptional children, administration of special education; the mentally handicapped; the mentally gifted; the socially maladjusted; the auditorily and the speech handicapped; the negroes; bilingual children; and the Indians.

International Bureau of Education and Post-war Educational Reconstruction. Bulletin No. 84. Geneva, Switzerland: International Bureau of Education, 1944. Pp. 7.

Discusses aid to educational and scientific libraries; school textbooks, foreign teaching staff; and post-war educational reforms.

International Breau of Education, 1942-1943. Bulletin No. 68. Geneva, Switzerland: International Bureau of Education, 1943. Pp. 37.

A summary of educational news from various foreign countries.

JAMES, T. L. *Standardizing Shorthand Tests.* Studies in Education, No. 10. Wellington, New Zealand: New Zealand Council for Educational Research, 1944. 64 pp.

Discusses the problem; examination procedure; linguistic factors; word-length; measurement of written shorthand; speaking-time-length; standard word measurement of past examinations; quarter-minute sections; and counting sounds.

KANDEL, I. L. *Intellectual Cooperation: National and International.* New York: Bureau of Publications, Teachers College, Columbia University, 1944. 78 pp.

Discusses education and international understanding; traditions of national education; need of a new spirit in education; humanism and international under-

standing; national mentalities and international cooperation; education of public opinion; specialization and compartmentalization or unification of knowledge; the place of science and technology in modern culture; learning and leadership; national intellectual cooperation; and organization of intellectual cooperation.

LOVE, JAMES LEE. *Reprint of a Harvard Report.* Published privately by author, 1944, Burlington, N. C. 18 pp.

This is a copy of the earliest written report to the President of the University by the officer in charge of the Harvard summer school of arts and sciences.

MCWILLIAMS, CAREY. *What About Our Japanese-Americans?* Public Affairs Pamphlet. New York City: Public Affairs Committee, Inc., 1944. Pp. 31.

Based on a comprehensive and fully documented study of the Japanese-Americans in peace and war.

PAQUIN, LAWRENCE. *Radio Program Production Aids.* Washington, D. C.: Federal Security Agency, U. S. Office of Education, 1944. 49 pp.

Presents suggestions to school and nonprofessional groups for the production of educational radio programs; a workbook of terms used in the production of radio programs; and the importance or radio sound effects.

POWERS, SAMUEL RALPH. *The Education of the Science Teacher.* Washington, D. C.: American Council of Science Teachers, National Education Association, 1942. 74 pp.

Discusses the function of science education in present-day America; the teacher as an individual and a citizen; the teacher as a person with proficiency in functional aspects of science; the science teacher as a person with proficiency in functional aspects of education; suggested organizations for education of the science teacher; and inservice education.

Preemployment Trainees and War Production. Vocational Division Bulletin No. 224, Defense Training Series No. 2. Washington, D. C.: Federal Security Agency, U. S. Office of Education, 1943. Pp. 88.

A report of a follow-up survey of persons receiving instruction in preemployment vocational training courses for war production workers.

Proceedings of the Institute of Health and Human Relations. Harrisburg, Pennsylvania: Division of Health Education, Department of Health, 1944. 57 pp.

A report of the fourth Health Institute held in Pennsylvania on the scientific angle of sex education. It discusses: life sciences as a background for health; life sciences in everyday living; basic factors in heredity; community organization for health education; the official agency; how sex is determined; sociological consequences of sex; how far should the school go in teaching sex; what is the teacher's share of responsibility for sex education; equitable divisions of responsibility in sex education problems—parent; and equitable divisions of responsibility in sex education problems—church.

Proceedings of the Ninth Annual Guidance Conference. Studies in Higher Education L. Lafayette, Indiana: Division of Educational Reference, Purdue University, 1944. 80 pp.

A discussion on guidance for youth; survey of a guidance program in a sec-

ondary school; pre-induction training implications for guidance; the vocational significance of the soldier's army experience; demobilization and readjustment of war workers and members of the armed forces; a sociographic study of a 10th-grade class; how society processes children; education and guidance of high-school youth in non-industrial communities; and science in the schools of tomorrow.

Proposals for Public Education in Post-war America. Washington, D. C.: National Education Association, 1944. Pp. 42.

Discusses effects of the war upon schools; desirable goals for public education after the war; training and readjustment of persons discharged from the armed forces or released from civilian occupations; personnel and cost of the suggested educational programs; and summary of needed action.

Recommendations of the International Conferences on Public Education. Bulletin No. 86. Geneva, Switzerland: International Bureau of Education, 1944. Pp. 43.

After having examined the results of the inquiries undertaken by the Bureau during the course of the year in collaboration with Ministries of Public Instruction, each of these International Conferences on public education adopted "recommendations" on the subjects studied, which were as follows: compulsory schooling and the raising of the school leaving age; admission to secondary schools; economies in the field of public education; the professional training of elementary school teachers; the professional training of secondary school teachers; councils of public instruction; the organization of special schools; the organization of rural schools; legislation regulating school buildings; school inspection; the teaching of modern languages; the teaching of psychology in the training of elementary and secondary school teachers; the salaries of elementary school teachers; the teaching of classical languages; the drafting, utilization and choice of school textbooks; the salaries of secondary school teachers; the organization of pre-school education; and the teaching of geography in secondary schools.

SEXSON, JOHN A. *Our Schools Fight for Liberty.* Annual Report, 1942–1943. Pasadena, California: Board of Education, 1944.

An illustrated report for the school year, 1942–43, of the Pasadena City High School District, the Pasadena City Junior College District, and the Pasadena City School District.

SIMMONS, KATHERINE. *The Brush Foundation Study of Child Growth and Development; II. Physical Growth and Development.* IX, No. 1. Washington, D. C.: Society for Research in Child Development, National Research Council, 1944.

Presents a discussion of selection criteria; cross-sectional age-sex constants for anthropometric dimensions; stature-weight-skeletal age intercorrelations; relative growth; stature and weight self-predictability and prediction of terminal stature; correlations of IQ with stature, with weight and with skeletal age; and selected references.

Research News and Communications

Address all research news and communications to Carter V. Good, Teachers College, University of Cincinnati, Cincinnati, Ohio.

Crucial Problems.—Florence Hale, editor of the *Grade Teacher,* outlines six critical issues of 1944:

1. Need for federal and state support of education, which is directly concerned with teacher supply and demand in wartime. This situation is fully as acute as last year.

2. Not only is the financial situation as regards teacher supply important for 1944, but a definite effort to impress teachers with the patriotic nature of their work is needed. The high pressure campaign among teachers for enlistment in the WACS, the WAVES, etc., has a glamorous appeal that continues to deplete our teacher supply.

3. A definite effort to have deferments made in the case of badly needed men. Each case should be studied definitely to make certain that the need is there, and then some way planned for indicating publicly the fact that the man teacher is not evading the draft, but has been definitely assigned to his position as a wartime service. Many men teachers hesitate to accept deferment, because there is no organized way of letting the public realize the situation.

4. Much publicity directed toward discouraging women with children under fifteen years of age from going into any employment (even teaching) to the detriment of home life and the future of the children. Child psychologists have given plenty of proof that, even though the working mother provides food and shelter and some sort of supervision, through hired help, yet great harm is often done to the children by the absence of the mother. She is often too weary, because of her work, or too greatly interested in outside affairs to give her little children the attention and sense of security so necessary to their proper development, especially as regards character.

5. Discouragement of a tendency to belittle the importance of cultural subjects in the school curriculum and to limit the instruction, particularly of children in rural areas, to the study of only those subjects concerned with securing a living in the particular community represented. This objective is important and has been neglected. The emphasis placed by some columnists and radio broadcasters on the impracticability of cultural subjects, or the study of other countries, is pernicious. Often it is so cleverly expressed that citizens do not think the matter through sufficiently. If we wish a real democracy, then we must not limit the education of our children to getting a living on a certain level that is to be made permanent. A community where the children have only a superficial knowledge of other countries, or where they have little of the cultural subjects, results in just the kind of a citizenship in which dictators rejoice, and for the elimination of which

this war is being fought. Children who are interested only slightly beyond their own limited environment become easy subjects for propaganda. A continued neglect of cultural subjects means a diminished ability to think logically and to express oneself adequately.

6. Not so much pre-occupation in post-war plans as to neglect present-day foundations. Post-war education is important, but until we have a better idea of when the war will end, the major amount of our time as regards elementary and secondary schools should be devoted to the immediate problems of keeping children in school, maintaining a safe and sane home atmosphere, and seeing that foundational work is done as complete as in normal times, even though as motivation it may be centered around war activities and future world building. In the long run, this will result in the best kind of post-war education. A thorough grounding in the regular school subjects will make a solid foundation upon which the special plans necessary to post-war work can be solidly built.

Extending Educational Opportunity of Children, Youth, and Adults.—The 1944 Yearbook of the Department of Supervision and Curriculum Development, N. E. A., has appeared under the title of *Toward a New Curriculum*. Many schools are initiating extensions of educational opportunities to new groups of people, in new areas of experience, and in new periods of the day and year. These practices grow out of developments in the current economic, political, and social scenes, in educational theory, and in changing concepts of the role of the school in the community. The concept that schooling includes more than reading, writing, and computing has been growing for many years. Practical realistic education has brought new patterns and new forms.

A number of publications of recent years have focused on the various facets of this developing pattern. By some these new activities were regarded as extra-curriculum, by others as passing fancies, and by still others as opportunistic educational activities, if not merely publicity stunts. In most cases the innovations have been unrecognized in their true light as signs of a changing curriculum. The authors of the yearbook view these changes as an extension of educational opportunity that is significant and promising for the future of American education. They believe that this extension of education represents a new approach in curriculum development.

The current innovations in the curriculum do not follow the orderly pattern of mass activity and assigned hours. Instead they disrupt the traditional program. Sometimes they are not even considered as being related to the school curriculum. Often they involve the hours before 9:00 and after 3:00. Saturdays, Sundays, holidays, and the usual summer vacation period may be the time when these activities flourish. While the program may be planned for learners in the six- to eighteen-year age range, it is possible that younger as well as older members of the community may be involved. These activities may take place in the school, but it is just as likely that they will not. Thus in more than one sense they represent a new development of the curriculum.

The very newness of these approaches means that the usual principles and practices of curriculum revision do not seem to fit. There are little reported experi-

ence and study by other groups that can be used as a basis for further planning. Real pioneering effort is necessary.

The authors of the yearbook face squarely the implications for curriculum development implied by these changes, and present a wealth of illustrations from school practice to show that extending educational opportunities to children, youth, and adults is no peripheral frill or fad, but is, in a real sense, a movement toward a new curriculum.

This summary was prepared by Ruth Cunningham, executive secretary of the Department of Supervision and Curriculum Development.

Research in Higher Education.—The recently published report *(Studies in Higher Education, 1940–42)* describes thirteen college investigations completed under the auspices of the University Committee on Educational Research at the University of Minnesota. As in the case of earlier volumes in this series, the investigations have been financed and coordinated by this committee, which has also furnished specialized statistical and consultative assistance, but the actual planning has been largely done by faculty members interested in studying their own curricular, personnel, and administrative problems.

Because of the many new responsibilities imposed on a university during war years, the resources of this committee during the past biennium have been devoted primarily to broad surveys of the courses offered throughout the university and of faculty services. Sub-committees in each of the colleges assisted in developing suitable schedules for obtaining the necessary data. Results from the curriculum study, based on detailed information concerning the four thousand courses currently offered by the University of Minnesota, and summarized by T. R. McConnell and Ruth E. Eckert, have been used to clarify aims of courses and curriculums and to identify undue overlapping among courses, as well as areas in which new courses should be developed in the post-war period. In the College of Education alone, the findings have constituted an important basis for dropping 28 courses, combining 66 courses into 31, and omitting another 16 courses. The university study of faculty load (reported by Ruth E. Eckert) revealed the many different kinds of contributions faculty members have been making to campus and community life. It showed, for example, that classroom teaching and preparation for such teaching normally claim only 42 per cent of the 55 hours that the typical faculty member in each of the colleges devotes to his work, the rest of his time being given to such frequently unrecognized services as student counseling, committee deliberations, administrative duties, advisement of professional and lay groups, and research and writing. It was also discovered that staff membership in one college imposed a somewhat different pattern of obligations than did membership in another division. On the basis of these findings the university has been enabled to use its staff more efficiently and equitably during the emergency and to negotiate sensible teaching loads for its military training programs.

Two of the remaining studies are university- or state-wide in scope, while the others deal with problems that have arisen in particular divisions or colleges. Included in the former group are an analysis of the residence and migration

of Minnesota college students, made in cooperation with the U. S. Office of Education and reported by Ruth E. Eckert, and a survey of the distribution of financial assistance to students in the several colleges of the University, by E. G. Williamson and John D. Foley. Important liberal-arts problems are considered in studies of the patterns of courses taken by majors in various fields who subsequently received the B. A. degree (T. R. McConnell and Ruth E. Eckert), the extent and character of student mortality (Kenneth E. Clark), factors related to success or failure in freshmen writing (Margaret E. Day), and the predictive values of measures commonly used for placing students in freshmen English courses (N. A. Fattu, James T. Hillhouse, and E. G. Williamson). Two of the investigations were made in the General College, and deal with the relation of study skills in the social studies to other measured-abilities (Horace T. Morse, George H. McCune, and Frances O. Triggs) and with the merits of socio-civic courses organized on a functional-need basis, as contrasted to those that adhere to the more conventional subject-matter approach (Horace T. Morse). The School of Business is represented by a study of measures predictive of success in its major curriculums (John G. Darley and Ralph F. Berdie), the School of Social Work by an analysis of the curriculum implications of a recent survey of the interests, activities and needs of social workers in Minnesota (Anne F. Fenlason and Raymond F. Sletto), and the College of Pharmacy by a project involving construction and validation of comprehensive examinations (Palmer O. Johnson).

While these are primarily service studies, made in an effort to develop educational policies suited to the peculiar needs of the students, faculty, and environing community of the University of Minnesota, many other faculties would undoubtedly find some of the techniques distinctly valuable in attacking analogous problems on their own campuses. Moreover, the notably successful attempts made at Minnesota to enlist the insights, enthusiasms, and talents of faculty members in all divisions in a program of continuous study and self-evaluation may well stimulate other colleges and universities to appraise more carefully their own procedures and practices.

This report was prepared by Ruth E. Eckert, associate director of the Committee on Educational Research.